

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:39:51 / Search time 235.2 Seconds
(without alignments)
3623.632 Million cell updates/sec

Title: US-09-889-325-4
Sequence: 1 MERLIDVERLQAMERAFRR.....HLSPHALVGLVATRELLQVAR 1208
6424

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: UniProt 05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	1 RECO4 HUMAN	094761 homo sapien
2	3944	61.4	1216	1 RECO4 MOUSE	075nr7 mus musculu
3	2859.5	41.4	1500	2 Q4YXK6 XENLA	04jnr8 xenopus lae
4	2310.5	36.0	910	2 Q4RLC3 XENLA	04jnr8 xenopus lae
5	1723.5	26.8	1579	2 Q9VSE6 DROME	09vns6 drosophila
6	1680.5	26.2	1530	2 Q9NH11 DROME	09nh11 drosophila
7	1670.5	26.0	1058	2 Q7PMM8 ANOCA	07pms8 anopheles g
8	1172	18.2	340	2 Q7YR85 BOVIN	07yrs5 bos taurus
9	971.5	15.1	927	2 Q7X829 ORYSA	07x829 oryza sativ
10	897.5	14.0	870	2 Q8W028 ARATH	08w028 arabidopsis
11	872	13.6	941	2 Q9C6N0 ARATH	09c6n0 arabidopsis
12	761	11.8	874	2 Q7FAH0 ORYSA	07fah0 oryza sativ
13	635	9.9	704	2 Q7MID7 VIBRI	07mid7 vibrio vuln
14	631.5	9.8	608	1 RECO EGOLI	08x81 escherichia
15	631.5	9.8	611	2 Q8X8T1 ECOS7	08x8t1 escherichia
16	631.5	9.8	611	2 Q8F8M6 ECOS6	08f8m6 escherichia
17	631.5	9.8	611	2 Q831V3 SHIFL	0831v3 shigella fi
18	626	9.7	639	2 Q8DB16 VIBRI	08db16 vibrio vuln
19	617.5	9.6	654	2 Q87MB5 VIBRI	087mb5 vibrio para
20	613.5	9.6	608	2 Q6CZH9 ERWCT	06czh9 erwinia car
21	613	9.5	642	2 Q87J96 PSEEM	087j96 pseudomonas
22	609.5	9.5	609	2 Q8ZB32 SALTI	08zb32 salmonella
23	609.5	9.5	615	2 Q57H99 SALCH	057h99 salmonella
24	607.5	9.5	608	1 RECO SALTY	08k128 pseudomonas
25	604	9.4	644	2 Q4KJ28 PSEFS	04k128 pseudomonas
26	604	9.4	649	2 Q4ZL17 PSESY	04zl17 pseudomonas
27	604	9.4	654	2 Q88EKL SHRON	088ekl shewanella
28	603.5	9.4	609	2 Q5PRK6 SALPA	05prk6 salmonella
29	599	9.3	746	2 Q7NH48 GLOVI	07nh48 gloeobacter
30	594.5	9.3	595	2 Q4NM16 JDDEL	04nm16 anaeromyob
31	593	9.2	610	2 Q66FY3 YERSIN	066fy3 yersinia ps

32	593	9.2	610	2 Q8ZAG8 YERPE	08zag8 yersinia pe
33	592.5	9.2	464	2 Q5DYC0 VIBF1	05dyc0 vibrio fisc
34	586.5	9.1	608	2 Q7MYQ5 PHOLI	07myq5 photorhabdu
35	586	9.1	602	2 Q88Y35 LACPL	088y35 lactobacill
36	584.5	9.1	641	2 Q6IK00 PHOPR	06ik00 photobacter
37	582.5	9.1	651	2 Q5FPA4 GLUCOB	05fpa4 glucobacter
38	581.5	9.1	620	2 Q9KVF0 VIBCH	09kvf0 vibrio chol
39	580.5	9.0	749	2 Q7UR05 RHOB	07ur05 rhodospirill
40	578.5	9.0	632	1 RECO PASMO	09c121 pasteurella
41	578	9.0	619	2 Q4QMG6 HAIR8	04qmg6 haemophilus
42	577	9.0	619	2 Q5IM08 SILPO	05im08 silicibacter
43	575	9.0	679	2 Q5LIP9 PHOPR	05lip9 photobacter
44	573.5	8.9	615	2 Q65QS2 MANSM	065qs2 mannheimia
45	573	8.9	637	2 Q65QS2 MANSM	065qs2 mannheimia

ALIGNMENTS

RESULT 1
ID RECO4 HUMAN STANDARD; PRT; 1208 AA.
AC 094761, Q96DM2, Q96P55;
AT 30-MAY-2000 (Rel. 39, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE ATP-dependent DNA helicase Q4 (EC 3.6.1.1) (RecQ protein-like 4)
DB (RecQ4) (RFS).
GN Name=RecQ4; Synonym=RECQ4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC TISSUE=Testis;
RX MEDLINE=99097344; PubMed=9878247; DOI=10.1006/geno.1998.5595;
RA Kitao S., Ohnogi I., Ichikawa K., Goto M., Furuchi Y., Shimamoto A.,
RT "Cloning of two new human helicase genes of the RecQ family:
RT biological significance of multiple species in higher eukaryotes."
RL Genomics 54:443-452(1998).
[2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], SUBCELLULAR LOCATION, AND DISEASE.
RX MEDLINE=20021764; PubMed=10552928; DOI=10.1006/geno.1999.5959;
RA Kitao S., Lindor N.M., Shiratori M., Furuchi Y., Shimamoto A.,
RT "Rothmund-Thomson syndrome responsible gene, RECQ4: genomic structure
RT and products."
RL Genomics 61:268-276(1999).
[3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 468-1208.
RC TISSUE=Lymph, and Placenta;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RT Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RT Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.T., Wang J., Heich F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M., Soares M.B., Bonaldo M.F., Casavant T.L., Senech T.E.,
RA Srapstein M.J., Udell T.B., Tothilyuk S., Carninci P., Prange C.,
RA Bha S.S., Loughran J.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bha S.S., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywiński M.I., Skalka U., Smallus D.B.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]
RP INDUCTION.
RX PubMed=11032027; DOI=10.1036/sj onc.1203841;
RA Kawabe T., Taniyama N., Kikao S., Nishikawa K., Shimamoto A.,
RA Shitatori M., Matsumoto T., Anno K., Sato T., Mitsui Y., Seki M.,
RA Konomoto T., Goto M., Ellis N.A., Ide T., Furutachi Y., Sugimoto M.,
RT "Differential regulation of human RecQ family helicases in cell
RT transformation and cell cycle."
RL Oncogene 19:4764-4772(2000).
[5]
RP DISEASE.
RX PubMed=12952869; DOI=10.1093/hmg/ddg306;
RA Siltonen H.A., Kopra O., Kaeaeilaenen H., Haravuori H., Winter R.M.,
RA Saesamaenen A.-M., Peltonen L., Keskihae M.,
RT "Molecular defect of RAPADILINO syndrome expands the phenotype
RT spectrum of RECQL diseases."
RL Hum. Mol. Genet. 12:2837-2844(2003).
[6]
RP INTERACTIONS WITH UBR1 AND UBR2, SUBCELLULAR LOCATION, FUNCTION, AND
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX PubMed=15317757; DOI=10.1093/hmg/ddh259;
RA Yin J., Kwon Y.T., Varshavsky A., Wang W.,
RA "RECQL4, mutated in the Rothmund-Thomson and RAPADILINO syndromes,
RT interacts with ubiquitin ligases UBR1 and UBR2 of the N-end rule
RT pathway."
RL Hum. Mol. Genet. 13:2421-2430(2004).
[7]
RP VARIANTS ASP-267 AND GLN-1005.
RX PubMed=12601557; DOI=10.1007/s100380300016;
RA Roverai G., Beghini A., Zamburino G., Paradisi M., Larizza L.,
RT "Identification of two novel RECQL4exonic SNPs and genomic
RT characterization of the IVS12 minisatellite."
RL J. Hum. Genet. 48:107-109(2003).
[8]
RP VARIANTS GLN-355; SER-441 AND 857-CYS--THR-858 DEL.
RX PubMed=15221963; DOI=10.1002/jic.20269;
RA Nishijo K., Nakayama T., Aoyama T., Okamoto T., Ishibe T., Yasura K.,
RA Shima Y., Shibata K.R., Tsuboyama T., Nakamura T., Tochida J.,
RT "Mutation analysis of the RECQL4 gene in sporadic osteosarcomas."
RL Int. J. Cancer 111:367-372(2004).
-1- FUNCTION: DNA-dependent ATPase. May modulate chromosome
segregation.
-1- SUBUNIT: Interacts with UBR1 and UBR2.
-1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
-1- TISSUE SPECIFICITY: Ubiquitously expressed, with highest levels in
thymus and testis.
-1- INDUCTION: Up-regulated in actively dividing cells.
-1- DISEASE: Defects in RECQL4 are a cause of Rothmund-Thomson
syndrome (RTS) [MIM:268400]. A disease characterized by
dermatological features such as atrophy, pigmentation, and
telangiectasia and frequently accompanied by juvenile cataract,
saddle nose, congenital bone defects, disturbances of hair growth,
and hypogonadism.
-1- DISEASE: Defects in RECQL4 are a cause of RAPADILINO syndrome
[MIM:266280]. A disease characterized by radial and patellar
aplasia or hypoplasia.
-1- SIMILARITY: Belongs to the helicase family. RecQ subfamily.
-1- DATABASE: NMB=Atlas Genet. Cytogenet. Oncol. Haematol.;
WWW="http://www.infobiogen.fr/services/chronocancer/Genes/RECQL4ID285.html"

CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.

CC EMBL; AB006532; BAA74453.1; -; mRNA.
DR EMBL; AB026546; BAA66899.1; -; Genomic_DNA.
DR EMBL; BC011602; AAH11602.2; -; mRNA.
DR EMBL; BC013277; AAH13277.2; ALT_INIT; mRNA.
DR HSSP; F15043; 10Y.
DR Ensemble; ENSG00000160957; Homo sapiens.
DR HGNC; HGNC:9949; RECQL4.

DR	MIM; 603780;	-
DR	MIM; 268400;	-
DR	MIM; 268280;	-
DR	GO; GO:0003678;	F:DNA helicase activity; TAS.
DR	GO; GO:00022275;	P:development; TAS.
DR	GO; GO:0006281;	P:DNA repair; TAS.
DR	InterPro; IPR001410;	DEAD.
DR	InterPro; IPR011545;	DEAD/DEAH_N.
DR	InterPro; IPR002464;	DEAH box.
DR	InterPro; IPR01650;	Helicase_C.
DR	InterPro; IPR004589;	RecQ.
DR	PANTHER; PTHR13710;	RecQ; 1.
DR	Pfam; PF00270;	DEAD; 1.
DR	Pfam; PF00271;	Helicase_C; 1.
DR	TIGRFAMs; TIGR00614;	recq_fam; 1.
DR	PROSITE; PS00690;	DEAH ATP Helicase; FALSE_NEG.
KW	ATP-binding; Cataract;	Helicase; Hydrolyase; Nuclear protein;
KW	Nucleotide-binding;	Polymorphism.
FT	NP_BIND	502 509 ATP (Potential).
FT	MOTIF	605 608 DEAH box. E -> D (common polymorphism).
FT	VARIANT	267 267 /FTID=VAR_023295. R -> Q.
FT	VARIANT	355 355 /FTID=VAR_023296. P -> S.
FT	VARIANT	441 441 /FTID=VAR_023297. Missing.
FT	VARIANT	857 858 /FTID=VAR_023298. R -> Q (common polymorphism).
FT	VARIANT	1005 1005 /FTID=VAR_023299.
FT	SEQUENCE	1208 AA; 133077 MW; CB809A7765AB48A1 CRC64;
Query Match	100.0%;	Score 6424; DB 1; Length 1208;
Bases Local Similarity	100.0%;	Pred. No. 2.7e-307;
Matches 1208;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0
OY	1	MERLADVVERIQAMERARFRORGRPPSODDYEAPEETRALVREYRTLLKRTTGOAGGRL 60
Db	1	MERLADVVERIQAMERARFRORGRPPSODDYEAPEETRALVREYRTLLKRTTGOAGGRL 60
OY	61	SSESLPAAEBAEPBRCWGPMLNRAATKSPOTPGRSRGSVPDYGORLKANIKGTLOAG 120
Db	61	SSESLPAAEBAEPBRCWGPMLNRAATKSPOTPGRSRGSVPDYGORLKANIKGTLOAG 120
OY	121	PALRRRPMPPLGRASSASKASTPKPGCGIPFSPFAEKVSDSPOLPEBPQPRGRIQHLOASIS 180
Db	121	PALRRRPMPPLGRASSASKASTPKPGCGIPFSPFAEKVSDSPOLPEBPQPRGRIQHLOASIS 180
OY	181	QRLSGLDPGMTORCHSEVFDFLGAPKACRPDLGSEESQLLYRGESAVLGRGSGSGPEAS 240
Db	181	QRLSGLDPGMTORCHSEVFDFLGAPKACRPDLGSEESQLLYRGESAVLGRGSGSGPEAS 240
OY	241	AFOEVSIIRVSGPOPSSSGGERRNNWEBSWESPAAVOQESSQAQPSESGAGAIAVEEDPG 300
Db	241	AFOGVSIIRVSGPOPSSSGGERRNNWEBSWESPAAVOQESSQAQPSESGAGAIAVEEDPG 300
OY	301	EPVVAOPPOPCCSSSNPRHYHGLSPSSQARAAGASTPLHIIPRLARDHRGNVYILNMKO 360
Db	301	EPVVAOPPOPCCSSSNPRHYHGLSPSSQARAAGASTPLHIIPRLARDHRGNVYILNMKO 360
OY	361	KHYVRGALNSRLRKQAMKOKMKKKGCFFGGGATVTTKSCFLNEGFDDHAACCPRA 420
Db	361	KHYVRGALNSRLRKQAMKOKMKKKGCFFGGGATVTTKSCFLNEGFDDHAACCPRA 420
OY	421	SEBDTDAVGPEPIVPSPOPVESVDPTVLPLYSLGPSGLAETPAVFAVOLBOIGQA 480
Db	421	SEBDTDAVGPEPIVPSPOPVESVDPTVLPLYSLGPSGLAETPAVFAVOLBOIGQA 480
OY	481	FRPQGEAWNRITISGITLVLPFGAGSGSLCYOYPALLYSRSPCLTLTVVSPTSLSMDQ 540
Db	481	FRPQGEAWNRITISGITLVLPFGAGSGSLCYOYPALLYSRSPCLTLTVVSPTSLSMDQ 540
OY	541	VSGIPLCKAKACHSGMTRKQRESVLEKIRANAQVHVLMLTBEALVGAGGLPPAAQLPVA 600

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	481	7.5	767	6	US-10-467-657-2430	Sequence 2430, Ap
2	481	7.5	767	6	US-10-467-657-6630	Sequence 6630, Ap
3	476	7.4	1405	6	US-10-995-561-529	Sequence 529, App
4	476	7.4	1406	6	US-10-995-561-530	Sequence 530, App
5	476	7.4	1436	6	US-10-995-561-531	Sequence 531, App
6	232.5	3.6	1823	6	US-10-995-561-988	Sequence 988, App
7	232.5	3.6	2102	6	US-10-995-561-990	Sequence 990, App
8	232.5	3.6	2157	6	US-10-995-561-991	Sequence 991, App
9	229	3.6	2108	6	US-10-995-561-989	Sequence 989, App
10	225.5	3.5	1335	6	US-10-793-626-3334	Sequence 3334, App
11	224	3.5	411	6	US-10-793-626-3156	Sequence 3156, Ap
12	207.5	3.2	614	6	US-10-519-447-2	Sequence 2, Appli
13	207.5	3.2	614	6	US-10-878-556A-75	Sequence 75, Appl
14	199	3.1	882	7	US-11-012-762-34	Sequence 34, Appl
15	194.5	3.0	650	6	US-10-878-556A-110	Sequence 110, App
16	191.5	3.0	457	6	US-10-467-657-2172	Sequence 2172, Ap
17	170.5	2.7	874	7	US-11-012-762-8	Sequence 8, Appli
18	169.5	2.6	411	6	US-10-821-234-1386	Sequence 1386, Ap
19	169.5	2.6	411	6	US-10-878-556A-140	Sequence 140, App
20	166	2.6	1560	7	US-11-059-983-1	Sequence 1, Appli
21	165	2.6	406	6	US-10-878-556A-41	Sequence 41, Appl
22	163	2.5	915	6	US-10-995-561-1003	Sequence 1003, Ap
23	163	2.5	940	6	US-10-995-561-1004	Sequence 1004, Ap
24	163	2.5	969	6	US-10-995-561-1001	Sequence 1001, Ap
25	163	2.5	994	6	US-10-995-561-997	Sequence 997, App


```
QY 1050 QICDFLYGR-VOAREROA 1066
Db 1157 QI--VLXGKLYEARQKHA 1172

RESULT 6
US-10-995-561-988
; Sequence 988, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 988
; LENGTH: 1823
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-988

Query Match 3.6%; Score 232.5; DB 6; Length 1823;
Best Local Similarity 22.6%; Pred. No. 1.7e-07;
Matches 22; Conservative 72; Mismatches 325; Indels 363; Gaps 47;

QY 15 ERAPRRQRGRPSODDVEAAPEETRALYREYRTLKRTTGQAGGLRSESLSLPA--EE 71
Db 904 ERPRRRHGRAQQD-----KPRFRRLKQE-----RENAARGSEKPSLTLPASAPGEE 954
QY 72 -----APEPRCWGPHLNRAATKSPQ-----
Db 955 ALTTVTVPAPR-----RAAAKSPDLNSQNSQANEWEETASESDFTSERRGDKPEAP 1007
QY 92 -----PTGRSRQGSVP-----DYQRLKANLKGTLQA-CPAL---GRPPWP 129
Db 1008 PVLTLTPKAVGTPGGGGGAVPGISAMSRGDLQSRADLSKRFSFSSQRPQGMQRNRRPGP 1067
QY 130 LGRASSKASTPKPGTGP-----VPSFAEKVS-----DEPPOLPEQP----- 167
Db 1068 GKGAGSSGSSGGGGGPGGRTGCGRDKRSWPKNRSRPPKRSRPPKRLPLPPPPSSA 1127
QY 168 -----RPGRLQHLQASLSQRLGSLD-PGWLQR-----
Db 1128 VFRLDQVIHNSNPAGIQQAQLSSRQGSVTPAGGHPKHPKPPQAPQGPSPRPPTRYEPQ 1187
QY 194 -----CHSEVPDFL-----GAPKAC-----RDLGSEESQLLIPG 223
Db 1188 RVNSGLSSDPHFEPGPMVRGVGTPRDSAGVSPFPKRRERPRKPELLOEES-LPPPH 1246
QY 224 ESAVLGAGSGQGEAFAQVSVTRVQSP-----QFSSSGGEKRWNERPWESE- 272
Db 1247 SSGFLGSKPEGPGQAES-RDTGTEALTPHWNRLHTATSKSRPSMSEPMPEPLSPFE 1305
QY 273 -----AQV-----QOESS-----QAGPPSEAG-----AVAVE 295
Db 1306 DVAGTEMSQSDSGVDLSDGSQVSSGPCSQRSRSPDGLKGAABGPKKPPKRGSSSPLNAVPE 1365
QY 296 EDPGGEVQAPQPPQFCSSPSN--PRYHGLSPS--SQARAGKAE-GTAPLHIFPRLARHDR 350
Db 1366 GPPGSEPPRPPAPHDGDKELPREQLPPGPIGTERSQRTDGTGPIRP---SHRP 1422
QY 351 GNYVRLNMKQ-----HYVRGALRSRLLRKQAWTKWKXKBCGCGGATVTKESCFLN 406
Db 1423 GPPVQFGTSKDSRLRLVWGDSLK-----AEKE-----LTASVTEAIPVS 1462
QY 407 EQDF--HWAQCPAPSE--DTDVAGPEPLVPSQPVPEVPSLDPTVLPLY-SIGPSSQL 462
Db 1463 RDWELLSAASAPQSKNLDSCHCVEPSSSGORLYPEV-----PYGSAGPSSQ 1513
```

```
QY 463 AETPAEVFQALEQLGHQAFRPGQ-----ERAVNRILSGI-----S 497
Db 1514 ISGAMDSQLHPNSG--GFRFGTSLHPYRSQPLYLPPGPAPPSALLSGVALKGQFLDFS 1571
QY 498 TLIV-----LPTGAGKSLCYQLPALLYSR--SP-----CLTLVVSPLLSLMDDOVSGLP 545
Db 1572 TMOATELGKLPAG---GVLYPPPSFLYSPAFCSPLPDTSLQVRODLPSPSDFYSTPLQ 1628
QY 546 PCLKAACIHSGMTRKQRESVLQIRAAQVHVLMLTPEA-----LVGAGGLPPAAQLPPEVA 600
Db 1629 PGQSGFLPSGAPAQ-----MLPMWDSQLPVVNFSGSLPPAPPPPPP 1672
QY 601 FACIDEAHCLSQSHNFRPCYLKVRERMGVHCFGLGTATATRTTASDVAQHLAAVE 660
Db 1673 LSLLPVGPALQPPSLAVRP-----PPAPATRVLPSPARPPPASLG 1712
QY 661 EPDLHGAPVPTNHLHSVMDRDTQALLTLQKRRFQNLDSII-----IYCNRR 711
Db 1713 RAEHLHPVELKPFQDYQKLSNLGGPGSSRTPPTGRSFGSLNSRLKATPSTYSGVFTQRV 1772
QY 712 DTERIAA---LLRTCLHAAWVP 730
Db 1773 DLYQASPPDALR-----WIP 1788
```

```
RESULT 7
US-10-995-561-990
; Sequence 990, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 990
; LENGTH: 2102
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-990
```

```
Query Match 3.6%; Score 232.5; DB 6; Length 2102;
Best Local Similarity 22.6%; Pred. No. 2e-07;
Matches 22; Conservative 72; Mismatches 325; Indels 363; Gaps 47;

QY 15 ERAPRRQRGRPSODDVEAAPEETRALYREYRTLKRTTGQAGGLRSESLSLPA--EE 71
Db 1183 ERPRRRHGRAQQD-----KPRFRRLKQE-----RENAARGSEKPSLTLPASAPGEE 1233
QY 72 -----APEPRCWGPHLNRAATKSPQ-----
Db 1234 ALTTVTVPAPR-----RAAAKSPDLNSQNSQANEWEETASESDFTSERRGDKPEAP 1286
QY 92 -----PTGRSRQGSVP-----DYQRLKANLKGTLQA-CPAL---GRPPWP 129
Db 1287 PVLTLTPKAVGTPGGGGGAVPGISAMSRGDLQSRADLSKRFSFSSQRPQGMQRNRRPGP 1346
QY 130 LGRASSKASTPKPGTGP-----VPSFAEKVS-----DEPPOLPEQP----- 167
Db 1347 GKGAGSSGSSGGGGGPGGRTGCGRDKRSWPKNRSRPPKRSRPPKRLPLPPPPSSA 1406
QY 168 -----RPGRLQHLQASLSQRLGSLD-PGWLQR-----
Db 1407 VFRLDQVIHNSNPAGIQQAQLSSRQGSVTPAGGHPKHPKPPQAPQGPSPRPPTRYEPQ 1466
QY 194 -----CHSEVPDFL-----GAPKAC-----RDLGSEESQLLIPG 223
Db 1467 RVNSGLSSDPHFEPGPMVRGVGTPRDSAGVSPFPKRRERPRKPELLOEES-LPPPH 1525
```

```
QY 224 ESAVLGAGAGQPEASAFQEVSVIRVQSP-----QPSSSGGKKRWNEEPWESP----- 272
DB 1526 SSGFLGSKPEGPGQAES-RDTGTALTPHINWRLHTATSRKSYRPSMPEPMEPLSPFE 1584
QY 273 -----AQV-----QOESS-----QAGPPSEAG-----AVAVE 295
DB 1585 DVAGTEMSQSDSGVDLSGDSQVSSGPCSQSPDGGLKGAAGPKRPGSSPLNAVPC 1644
QY 296 EDPGEPVQAPPOPCSSPN--PRYHGLSPS--SQARAGKAE--GTAPLHIFPLARHDR 350
DB 1645 GPPGSEPPRRPPAPHGDRKELPREQPLPGPIGTERTSORDTGTGTPGPIRP---SHRP 1701
QY 351 GNYVRLNMKQK----HYVRGALRSRLLRKQAWKQKWKKECGFGGATVTTKESCFLN 406
DB 1702 GPPVQFGTSDKSDRLRLVVGDSLK-----AEKE-----LTASVTEAIPVS 1741
QY 407 EQFD--HWAACQPRPASEE--DITDAVGPEPLVPSQPVPVEVPSLDPVLPVLY--SLGPSQOL 462
DB 1742 RDWELLPSAASAEFQSKNLDGCHVPEPSSGQRLYPEV-----FYGSAGPSSQ 1792
QY 463 AETPAEVQFQALEQGHQAFRPGQ-----ERAVMRILSGI-----S 497
DB 1793 ISGAMDSQLHPNSG--GFRGTPSLHPYRSQPLYPGPPAPPALLSGVALKGQFLDPS 1850
QY 498 TLLV-----LPTGAGKSLCYQLPALLYRR--SP-----CLTVVSPLLSLMDDOVSGLP 545
DB 1851 TMQATELGKLPAG---GVLYPPPSFLYSPAFPCSPPLDTSLLQVRQDLPSDFYSTPLQ 1907
QY 546 PCLKAACTHSGWTRKQRESVLQIRAAQVHVLMLTPEA-----LVGAGGLPAAQLPPVA 600
DB 1908 PGGSGGLPSPGAPAQO-----MLLPWVDSQLPVNFSGSLPPAPPAPP 1951
QY 601 PACIDEAHCLSQWSHNFRPCYLVRCKVLRMGVHCFGLGTATATRTTASDVQAHLVAE 660
DB 1952 LSLLPVGPALQPPSLAVRP-----PPAPATRVLPSPARPPASLG 1991
QY 661 EPDLHGAPVPTNLHLSVMDRDTQALLTLQKRFQNLDSII-----IYCNRE 711
DB 1992 RAEHLFVELKPFQDYQKLSNLGGPSSRTPTGSRFSGLSRLKATPSTYSGVFTQRV 2051
QY 712 DTERIAA---LLRTCLHAATVP 730
DB 2052 DLYQOASPPDALR-----WIP 2067

RESULT 8
US-10-995-561-991
; Sequence 991, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 991
; LENGTH: 2157
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-991

Query Match 3.6%; Score 232.5; DB 6; Length 2157;
Best Local Similarity 22.6%; Pred. No. 2.1e-07;
Matches 222; Conservative 72; Mismatches 325; Indels 363; Gaps 47;

QY 15 ERAFRQRGRSQQDVAAPEETRALYREYRTLKTTGQAGGLRSSBSLPAQA---EE 71
DB 1238 ERPRRRHGAQQDD-----KPRFRRLKQF-----RENARSGKPSLTLPASAPGEE 1288
```

```
QY 72 -----APEPRCWGPHLNRATKSPQ----- 91
DB 1289 ALTTVTVPAPR-----RAAKSPDLNSQANEWEETASBSSDFTSERRGDEAP 1341
QY 92 -----PTFGRSRQGSVP-----DYQORLKANLKGTLQA--GPAL-----GRPPWP 129
DB 1342 PPVLLTPKAVGTTPGGGGGAVPGISAMSRGDLQSRANDLSKRFSQORPQMERQNRPPG 1401
QY 130 LGRASSKASTPKPPTGTP-----VPSFAEKYS-----DEPQLPEPQ----- 167
DB 1402 GKGAGSGSGSGGGGPGGRTGFGRGDKRSWSPKMRSPRPPPEPPLGLPLPPPPSSSA 1461
QY 168 -----RPGRLQHLQASLSORLGLD--PGWLQR----- 193
DB 1462 VFRLDQVIHNAPAGIQALQALSROGSVTAPGCHPHKPGPQAPQPGSPRPTRYPEQ 1521
QY 194 -----CHSEVPDFL-----GAPKAC-----RPDLSEESQLLIPG 223
DB 1522 RVNSGLSSDPHFEEPGVMVRGVGTPRDSAGVPPPKRRRPPRPRKPELLQEE--LPPPH 1580
QY 224 ESAVLGAGAGQPEASAFQEVSVIRVQSP-----QPSSSGGKKRWNEEPWESP----- 272
DB 1581 SSGFLGSKPEGPGQAES-RDTGTALTPHINWRLHTATSRKSYRPSMPEPMEPLSPFE 1639
QY 273 -----AQV-----QOESS-----QAGPPSEAG-----AVAVE 295
DB 1640 DVAGTEMSQSDSGVDLSGDSQVSSGPCSQSPDGGLKGAAGPKRPGSSPLNAVPC 1699
QY 296 EDPGEPVQAPPOPCSSPN--PRYHGLSPS--SQARAGKAE--GTAPLHIFPLARHDR 350
DB 1700 GPPGSEPPRRPPAPHGDRKELPREQPLPGPIGTERTSORDTGTGTPGPIRP---SHRP 1756
QY 351 GNYVRLNMKQK----HYVRGALRSRLLRKQAWKQKWKKECGFGGATVTTKESCFLN 406
DB 1757 GPPVQFGTSDKSDRLRLVVGDSLK-----AEKE-----LTASVTEAIPVS 1796
QY 407 EQFD--HWAACQPRPASEE--DITDAVGPEPLVPSQPVPVEVPSLDPVLPVLY--SLGPSQOL 462
DB 1797 RDWELLPSAASAEFQSKNLDGCHVPEPSSGQRLYPEV-----FYGSAGPSSQ 1847
QY 463 AETPAEVQFQALEQGHQAFRPGQ-----ERAVMRILSGI-----S 497
DB 1848 ISGAMDSQLHPNSG--GFRGTPSLHPYRSQPLYPGPPAPPALLSGVALKGQFLDPS 1905
QY 498 TLLV-----LPTGAGKSLCYQLPALLYRR--SP-----CLTVVSPLLSLMDDOVSGLP 545
DB 1906 TMQATELGKLPAG---GVLYPPPSFLYSPAFPCSPPLDTSLLQVRQDLPSDFYSTPLQ 1962
QY 546 PCLKAACTHSGWTRKQRESVLQIRAAQVHVLMLTPEA-----LVGAGGLPAAQLPPVA 600
DB 1963 PGGSGGLPSPGAPAQO-----MLLPWVDSQLPVNFSGSLPPAPPAPP 2006
QY 601 PACIDEAHCLSQWSHNFRPCYLVRCKVLRMGVHCFGLGTATATRTTASDVQAHLVAE 660
DB 2007 LSLLPVGPALQPPSLAVRP-----PPAPATRVLPSPARPPASLG 2046
QY 661 EPDLHGAPVPTNLHLSVMDRDTQALLTLQKRFQNLDSII-----IYCNRE 711
DB 2047 RAEHLFVELKPFQDYQKLSNLGGPSSRTPTGSRFSGLSRLKATPSTYSGVFTQRV 2106
QY 712 DTERIAA---LLRTCLHAATVP 730
DB 2107 DLYQOASPPDALR-----WIP 2122
```

RESULT 9

```
US-10-995-561-989
; Sequence 989, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
```

```
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C0001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 989
; LENGTH: 2108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-989

Query Match      3.6%; Score 229; DB 6; Length 2108;
Best Local Similarity 23.6%; Pred. No. 3.4e-07;
Matches 198; Conservative 56; Mismatches 258; Indels 328; Gaps 44;

QY 15 ERAPRRGRGRPSDDVEAAPEETRALVREYTLKRTTGGAGGGLRSSESIPAAA---EE 71
Db 1183 ERPRRRHGRAQQD-----KPRFRRLKQE-----RENAAGSGKPSLTLPASAPGPEE 1233
QY 72 -----APRRCWGPHLNRAATKSPQ-----
Db 1234 ALTTVTVPAPR-----RAAAKSPDLNQNQSDQANEWEETASESSDFTSERRGDKPEAP 1286
QY 92 -----PTPGRSROGVP-----DYQRLKANLKGTLQA-GPAL---GRRPWP 129
Db 1287 PVLLTPKAVGTPGGGGGAVPGISAMSRGDLQRAKDLKSRFSQRPQGMERONRRPFGP 1346
QY 130 LGRASSKASTPKPGTGP---VSPFAEKVS-----DEPQLPEPQP----- 167
Db 1347 GKGAGSGSSGGGGGGPGGTGPRGDKRSPWPKNRSRPPERRPGLPLPPPPSSSA 1406
QY 168 -----RPGRLQHLQASLQSLD-PGWLQR-----
Db 1407 VFRLDQVTHSNFAGIQALASLQSGSVTAPGCHPRKPGPPQAPQPGSPRPTRVEPQ 1466
QY 194 -----CHSEVPDFL-----CAPKAC-----RPLGSEESQLLIPG 223
Db 1467 RVNSGLSDPHFEPPGPMVRGVGTGRDSDAGVSPFPKRRRPRPKPELLQEEES-LPPPH 1525
QY 224 ESAILVPGAGSQGPEASAFQSVSRVSGP-----QSSSGGKRRWRNEPWPESP----- 272
Db 1526 SSGFLGSKPEPGQAQES-RDTGTEALTPIWNRLHTATRSKSVRPSMFMPEPLSPFE 1584
QY 273 -----AQV-----QBSSS-----QAGPPSPGAG-----AVAVE 295
Db 1585 DVAGTEMSSQSDGVDLSQSVSSGPCSQRSSPDGLKGAAGPPKPPGSSSPLNAVPC 1644
QY 296 EDPGEPVQAQPPQPCSSPSN--PRYHGLSPS--SQARAGKAE-GTAPLHIFPPLARHDR 350
Db 1645 GPPGSEPPRPPPPAPHDGRKELPREQPLPFGPIGTERTSQRDRTGTFPGPIRP---SHRP 1701
QY 351 GNYVRLNMKOK---HYVRGRALSRLLRQAWKQKWRKGEFCGGGATVTTKESCFLN 406
Db 1702 GPPVQFGTSKDSLDRLVVGDSLK-----AEKE-----LTASVTTEALPVS 1741
QY 407 EQFD--HMAAQCPRASEE-DTDVAGPEPLVPSPQVPPEVPSLDPTVLPY-SLGPSSQGL 462
Db 1742 RDWELPSSAAASAEFPQSKNLDGHCVEPSSGQRLYPEV-----FVSGAGPSSSQ 1792
QY 463 AETPAEVFOALEQLGHQAFRPQ-----
Db 1793 ISGGMDSQLHPNSG--GFRPFGTSLHPYRSQPLYLPFGPAPPSPALLSGVALKGQFLDFS 1850
QY 498 TLLV-----LPTGAGKSLCYQLPALLYRR--SP-----CLTLVVSPLLSLMDQDVSLP 545
Db 1851 TMOATELGKLPAG--GVLTPPPSFLSPAFPCPSPLDPTSLQVRQDLSPSPSYSTPLQ 1907
QY 546 PCLKAACHSGMTKQRESVLOKIRAAQVHVLMTPEA-----LVGAGGLPPAAQLPPVA 600
Db 1908 FGGQSGFLPSGAPAAQ-----MLLPWVDSQLPVVNFSGSLPPA---PPPA 1948
```

```
RESULT 10
US-10-793-626-3334
; Sequence 3334, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3334
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3334

Query Match      3.5%; Score 225.5; DB 6; Length 135;
Best Local Similarity 37.1%; Pred. No. 2.2e-08;
Matches 53; Conservative 30; Mismatches 47; Indels 13; Gaps 4;

QY 462 LASTPAEVFOALEQLGHQAFRPQGERAVMVRILSGISTLLVLTGAGKSLCYQLPALLYSR 521
Db 1 MQETLSHYF-----GYKSPRQGEIITKILNHOHTLGVLTGGGKSCICYQVPLMQGG 54
QY 522 RSPCLTLVVSPLLSLMDQVSGLP--LKAACHSGMTKQRESVLOKIRAAQVHVLMT 580
Db 55 -----TTIVISPLISLKMKDQVDQLQAMGIQAAYLNSLLTHKQKEIEBQIKRGALQFLVA 110
QY 581 PEALVGAG--GLPPAAQLPPVAF 601
Db 111 PERFFNTFLNLLRKIEIPLIAF 133

RESULT 11
US-10-793-626-3156
; Sequence 3156, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3156
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3156

Query Match      3.5%; Score 224; DB 6; Length 411;
Best Local Similarity 38.7%; Pred. No. 1e-07;
Matches 48; Conservative 19; Mismatches 43; Indels 14; Gaps 2;

QY 704 IYCNREDTERIAALRTCLHAAWPGSGGRAPKTTAEAYHAGCMCRRRRRVQRAFMQG 763
Db 48 IYCVSTRKQVEE-----LHEAL-----NSEKIKSTIYHAGLTNKRIEAQNDFLYD 93
QY 764 QLVVVVATVAFGMLDRPDVRAVLHLGLPPSFYSYVQAVCRAGDGGQPAHCHLFLPQGE 823
```

Db 94 RVEVVIATNAPGMDKSNRVYIHYNMPGDLESYYQAGRAGRDGLKSECLILFSERDK 153
 Qy 824 DLRE 827
 Db 154 GLHE 157

RESULT 12
 US-10-519-447-2
 ; Sequence 2, Application US/10519447
 ; Publication No. US20050244829A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
 ; APPLICANT: Makoto OGINO
 ; APPLICANT: Hideki ENDOH
 ; TITLE OF INVENTION: METHOD FOR SCREENING AN AGENT FOR IMPROVING INSULIN RESISTANCE
 ; FILE REFERENCE: Q85576
 ; CURRENT APPLICATION NUMBER: US/10/519,447
 ; CURRENT FILING DATE: 2004-12-30
 ; PRIOR APPLICATION NUMBER: PCT/JP03/08367
 ; PRIOR FILING DATE: 2003-07-01
 ; PRIOR APPLICATION NUMBER: JP 2002-193814
 ; PRIOR FILING DATE: 2002-07-02
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 614
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-519-447-2

Query Match 3.2%; Score 207.5; DB 6; Length 614;
 Best Local Similarity 22.6%; Pred. No. 1.9e-06;
 Matches 110; Conservative 63; Mismatches 222; Indels 91; Gaps 19;
 Qy 387 GECFGGGATVTK-----ESCFLNQFDHAAQCPRPASEEDTAVGPEPLVPS 436
 Db 31 GKFGNPGKLVKKKNWLDLPKFKFYQHPD-----LARRAQEVETYSRSEKITVRG 86
 Qy 437 ---POPVEPVSPLDPTVLYSLGSPGQLAETPAEVFOALQHQAFRPGQERAVMRIL 493
 Db 87 HNCPE-----VLFNFE-----ANFPANVMDVIARQNFTPTAIQAQGWVAL 129
 Qy 494 SGISTLLVPTGAGKSLCYQLPALLYRRSP-----CLTLVVSPLSLMDQVSG- 543
 Db 130 SGLDMVGVAQTGSGKTLGYLLPAIVHINHQPFLERGDGPICLVLAPTRELAQQVQVAAE 189
 Qy 544 -LPPC-LKAACTHSGMTRKQRESVLQIRAAQVHVLMTPEALVG--AGGLPPAAQLPP 598
 Db 190 YCRACRLKSTCIYGAPKGPQIRDLER-----GVETCIATPGLIDFLECG----KTNLRR 241
 Qy 599 VAFACIDEAHCLSQSHNFRPCYLKRVCKLRRMGVHCFGLGTATATRTTASDVAQ---H 655
 Db 242 TTYLVLEADRLDM--GFEPQIRKIVDQIRPDRQTLKWSATWPKVEVQLAEDFLKDYIH 299
 Qy 656 LAVAEEDPLHGPAPVPTNLHL-----SVSMRDRTDQALLTLQKKRFQNLDSIIICNRR 711
 Db 300 INI-----GALELSANHNLIQIVDVCHDVEKDEKLRLMEEIMSEKENKTIVFV--E 349
 Qy 712 DTERIAALLRCLHAAPVPGSGRAPKTTAEAYHAGMCSRERRRVRQAFMQGLRVVAT 771
 Db 350 TKRRCDLTRKMRDGM--PAMG-----IHGDKSQQERDQVNLNFKHGKAPILIAT 398
 Qy 772 VAFGMGLDRPDVRAVLHGLPSPESYVQAVGRAGRDGQPAHCHLFLOP-----QGEDLRE 827
 Db 399 DVASRGLEDVKFVINYDYNSSSEDIYHRIGRSTARSTKTGTAYTFTFPNNIKQVSDLS 458
 Qy 828 LRRHVH 833
 Db 459 VLREAN 464
 RESULT 13
 US-11-012-762-34
 ; Sequence 34, Application US/11012762
 ; Publication No. US20050244815A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Georgia State University Research Foundation, Inc.
 ; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
 ; FILE REFERENCE: GSU1.PCT
 ; CURRENT APPLICATION NUMBER: US/11/012,762
 ; CURRENT FILING DATE: 2004-12-15
 ; PRIOR APPLICATION NUMBER: PCT/US03/19300
 ; PRIOR FILING DATE: 2003-06-19

US-10-878-556A-75
 ; Sequence 75, Application US/10878556A
 ; Publication No. US20050286399A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hoffmann La-Roche Inc.
 ; TITLE OF INVENTION: HCV regulated protein expression
 ; FILE REFERENCE: 21762
 ; CURRENT APPLICATION NUMBER: US/10/878,556A
 ; CURRENT FILING DATE: 2004-06-28
 ; NUMBER OF SEQ ID NOS: 199
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 75
 ; LENGTH: 614
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: sw_hum/ddx5_human
 ; DATABASE ENTRY DATE: 1990-08-01
 US-10-878-556A-75

Query Match 3.2%; Score 207.5; DB 6; Length 614;
 Best Local Similarity 22.6%; Pred. No. 1.9e-06;
 Matches 110; Conservative 63; Mismatches 222; Indels 91; Gaps 19;
 Qy 387 GECFGGGATVTK-----ESCFLNQFDHAAQCPRPASEEDTAVGPEPLVPS 436
 Db 31 GKFGNPGKLVKKKNWLDLPKFKFYQHPD-----LARRAQEVETYSRSEKITVRG 86
 Qy 437 ---POPVEPVSPLDPTVLYSLGSPGQLAETPAEVFOALQHQAFRPGQERAVMRIL 493
 Db 87 HNCPE-----VLFNFE-----ANFPANVMDVIARQNFTPTAIQAQGWVAL 129
 Qy 494 SGISTLLVPTGAGKSLCYQLPALLYRRSP-----CLTLVVSPLSLMDQVSG- 543
 Db 130 SGLDMVGVAQTGSGKTLGYLLPAIVHINHQPFLERGDGPICLVLAPTRELAQQVQVAAE 189
 Qy 544 -LPPC-LKAACTHSGMTRKQRESVLQIRAAQVHVLMTPEALVG--AGGLPPAAQLPP 598
 Db 190 YCRACRLKSTCIYGAPKGPQIRDLER-----GVETCIATPGLIDFLECG----KTNLRR 241
 Qy 599 VAFACIDEAHCLSQSHNFRPCYLKRVCKLRRMGVHCFGLGTATATRTTASDVAQ---H 655
 Db 242 TTYLVLEADRLDM--GFEPQIRKIVDQIRPDRQTLKWSATWPKVEVQLAEDFLKDYIH 299
 Qy 656 LAVAEEDPLHGPAPVPTNLHL-----SVSMRDRTDQALLTLQKKRFQNLDSIIICNRR 711
 Db 300 INI-----GALELSANHNLIQIVDVCHDVEKDEKLRLMEEIMSEKENKTIVFV--E 349
 Qy 712 DTERIAALLRCLHAAPVPGSGRAPKTTAEAYHAGMCSRERRRVRQAFMQGLRVVAT 771
 Db 350 TKRRCDLTRKMRDGM--PAMG-----IHGDKSQQERDQVNLNFKHGKAPILIAT 398
 Qy 772 VAFGMGLDRPDVRAVLHGLPSPESYVQAVGRAGRDGQPAHCHLFLOP-----QGEDLRE 827
 Db 399 DVASRGLEDVKFVINYDYNSSSEDIYHRIGRSTARSTKTGTAYTFTFPNNIKQVSDLS 458
 Qy 828 LRRHVH 833
 Db 459 VLREAN 464

RESULT 14
 US-11-012-762-34
 ; Sequence 34, Application US/11012762
 ; Publication No. US20050244815A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Georgia State University Research Foundation, Inc.
 ; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
 ; FILE REFERENCE: GSU1.PCT
 ; CURRENT APPLICATION NUMBER: US/11/012,762
 ; CURRENT FILING DATE: 2004-12-15
 ; PRIOR APPLICATION NUMBER: PCT/US03/19300
 ; PRIOR FILING DATE: 2003-06-19

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:39:51 ; Search time 235.2 Seconds

(without alignments)
3623.632 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLRDVRERLQWERAFFR.....HLSEFHALVGLATELLOVAR 1208

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05_80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6424	100.0	1208	1	RECQ4_HUMAN	O94761 homo sapien
2	3944	61.4	1216	1	RECQ4_MOUSE	Q75nr7 mus musculus
3	2659.5	41.4	1500	2	Q4JNX8_XENLA	Q4jnx8 xenopus lae
4	2310.5	36.0	910	2	Q4RLC3_TETNG	Q4rlc3 tetraodon n
5	1723.5	26.8	1579	2	Q9VSE6_DROME	Q9vse6 drosophila
6	1680.5	26.2	1530	2	Q9NH11_DROME	Q9nh11 drosophila
7	1670.5	26.0	1058	2	Q7PNM8_ANOGA	Q7pnm8 anopheles g
8	1172	18.2	340	2	Q7YR85_BOVIN	Q7yr85 bos taurus
9	897.5	15.1	927	2	Q7X829_ORYZA	Q7x829 oryza sativ
10	871.5	14.0	870	2	Q8W028_ARATH	Q8w028 arabidopsis
11	872	13.6	941	2	Q9C6N0_ARATH	Q9c6n0 arabidopsis
12	761	11.8	874	2	Q7FAH0_ORYZA	Q7fah0 oryza sativ
13	635	9.9	704	2	Q7MID7_VIBVY	Q7mid7 vibrio vuln
14	631.5	9.8	608	1	RECQ_ECOLI	P15043 escherichia
15	631.5	9.8	611	2	Q8X8N1_ECO57	Q8x8n1 escherichia
16	631.5	9.8	611	2	Q8FBM6_ECOL6	Q8fbm6 escherichia
17	631.5	9.8	611	2	Q83IW3_SHIFL	Q83iw3 shigella fl
18	626	9.7	639	2	Q8DBIC_VIBVU	Q8dbic vibrio vuln
19	617.5	9.6	654	2	Q7MBS5_VIBPA	Q7mbs5 vibrio para
20	613.5	9.6	608	2	Q6CZJ9_ERWCT	Q6czj9 erwinia car
21	613	9.5	642	2	Q87U96_PSESM	Q87u96 pseudomonas
22	609.5	9.5	609	2	Q823B2_SALTI	Q823b2 salmonella
23	609.5	9.5	615	2	Q57HP9_SALCH	Q57hp9 salmonella
24	607.5	9.5	608	1	RECQ_SALTY	P40724 salmonella
25	607	9.4	644	2	Q4KJZ8_PSEF5	Q4kjjz pseudomonas
26	604	9.4	649	2	Q4ZLJ7_PSESY	Q4zlj7 pseudomonas
27	604	9.4	654	2	Q8EKL1_SHEON	Q8eek1 shewanella
28	603.5	9.4	609	2	Q5PKN6_SALPA	Q5pkn6 salmonella
29	599	9.3	746	2	Q7NH48_GLOVI	Q7nh48 gloeobacter
30	594.5	9.3	595	2	Q4NW16_9DELT	Q4nw16 anaeromyxob
31	593	9.2	610	2	Q66FY3_YERSIN	Q66fy3 yersinia ps

32	593	9.2	610	2	Q8ZAG8_YERPE	Q8zag8 yersinia pe
33	592.5	9.2	464	2	Q5DYC0_VIBF1	Q5dyco vibrio fisc
34	586.5	9.1	608	2	Q7MYQ5_PHOLL	Q7myq5 photorhabdu
35	586	9.1	602	2	Q88Y35_LACPL	Q88y35 lactobacill
36	584.5	9.1	641	2	Q6LK00_PHOPR	Q6lk00 photobacter
37	582.5	9.1	651	2	Q5FNA4_GLUOX	Q5fna4 gluconobact
38	581.5	9.1	620	2	Q9KVF0_VIBCH	Q9kvf0 vibrio chol
39	580.5	9.0	749	2	Q7UR05_RHOBA	Q7ur05 rhodopirell
40	578.5	9.0	632	1	RECQ_PASMU	Q9cl21 pasteuralla
41	578	9.0	619	1	RECQ_HAEIN	P71339 haemophilus
42	577	9.0	619	2	Q4QMG6_HAEI8	Q4qmg6 haemophilus
43	575	9.0	679	2	Q5LWQ8_SILPO	Q5lwg8 silicibacte
44	573.5	8.9	615	2	Q6LLP9_PHOPR	Q6llp9 photobacter
45	573	8.9	637	2	Q65QS2_MANSM	Q65qs2 mannheimia

ALIGNMENTS

RESULT 1		STANDARD;		PRT; 1208 AA.	
RECQ4_HUMAN		AC	O94761; Q96DW2; Q96F55;		
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	13-SEP-2005 (Rel. 48, Last annotation update)				
DE	ATP-dependent DNA helicase Q4 (EC 3.6.1.-) (RecQ protein-like 4)				
DE	(RecQ4) (RTS).				
GN	Name=RECQ4; Synonym=RECQ4;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;				
OC	Homo.				
OX	NCBI_TaxID=9606;				
[1]					
RP	NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.				
RC	TISSUE=Testis;				
RX	MEDLINE=99097344; PubMed=9878247; DOI=10.1006/geno.1998.5595;				
RA	Kitao S., Ohsugi I., Ichikawa K., Goto M., Furuichi Y., Shimamoto A.;				
RT	"Cloning of two new human helicase genes of the RecQ family:				
RT	biological significance of multiple species in higher eukaryotes."				
RL	Genomics 54:443-452(1998). <				
[2]					
RN	NUCLEOTIDE SEQUENCE [GENOMIC DNA], SUBCELLULAR LOCATION, AND DISEASE.				
RP	MEDLINE=20021764; PubMed=10552928; DOI=10.1006/geno.1999.5959;				
RA	Kitao S., Lindor N.M., Shiratori M., Furuichi Y., Shimamoto A.;				
RT	"Rothmund-Thomson syndrome responsible gene, RECQL4: genomic structure				
RT	and products."				
RL	Genomics 61:268-276(1999).				
[3]					
RN	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 468-1208.				
RP	TISSUE=Lymph, and Placenta;				
RC	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;				
RX	Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,				
RA	Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,				
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,				
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,				
RA	Rana S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,				
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,				
RA	Richards S., Wozny K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,				
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,				
RA	Pabey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,				
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,				
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,				
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,				
RA	Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smalhus D.E.,				
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;				
RT	"Generation and initial analysis of more than 15,000 full-length human				
RT	and mouse cDNA sequences."				
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).				

[4]
RN INDUCTION.
RX PubMed=11032027; DOI=10.1038/sj.onc.1203841;
RA Kawabe T., Tsuyama N., Kitao S., Nishikawa K., Shimamoto A.,
RA Shiratori M., Matsumoto T., Anno K., Sato T., Mitsu Y., Seki M.,
RA Enomoto T., Goto M., Ellis N.A., Ide T., Furuchi Y., Sugimoto M.;
RT "Differential regulation of human RecQ family helicases in cell
transformation and cell cycle.";
RL Oncogene 19:4764-4772(2000).
RN [5]
RX PubMed=12952869; DOI=10.1093/hmg/ddq306;
RA Siitonen H.A., Kopra O., Kaeerlahti H., Haravuori H., Winter R.M.,
RA Saenmanen A.-M., Peltonen L., Kestilä M.;
RT "Molecular defect of RAPADILINO syndrome expands the phenotype
spectrum of RECQL diseases.";
RL Hum. Mol. Genet. 12:2837-2844(2003).
RN [6]
RX INTERACTIONS WITH UBR1 AND UBR2, SUBCELLULAR LOCATION, FUNCTION, AND
IDENTIFICATION BY MASS SPECTROMETRY.
RX PubMed=15317757; DOI=10.1093/hmg/ddb269;
RA Yin J., Kwon Y.T., Varshavsky A., Wang W.;
RT "RECQL4, mutated in the Rothmund-Thomson and RAPADILINO syndromes,
interacts with ubiquitin ligases UBR1 and UBR2 of the N-end rule
pathway.";
RL Hum. Mol. Genet. 13:2421-2430(2004).
RN [7]
RX VARIANTS ASP-267 AND GLN-1005.
RX PubMed=12601557; DOI=10.1007/s100380300016;
RA Roversi G., Beghini A., Zambruno G., Paradisi M., Larizza L.;
RT "Identification of two novel RECQL4 exonic SNPs and genomic
characterization of the IVS12 minisatellite.";
RL J. Hum. Genet. 48:107-109(2003).
RN [8]
RX VARIANTS GLN-355, SER-441 AND 857-CYS--THR-858 DEL.
RX PubMed=15221963; DOI=10.1002/ijc.20269;
RA Nishiyo K., Nakayama T., Aoyama T., Okamoto T., Ishibe T., Yasura K.,
RA Shima Y., Shibata K.R., Teuboyama T., Nakamura T., Toguchida J.;
RT "Mutation analysis of the RECQL4 gene in sporadic osteosarcomas.";
RL Int. J. Cancer 111:367-372(2004).
CC -1- FUNCTION: DNA-dependent ATPase. May modulate chromosome
segregation.
CC -1- SUBUNIT: Interacts with UBR1 and UBR2.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed, with highest levels in
thymus and testis.
CC -1- INDUCTION: Up-regulated in actively dividing cells.
CC -1- DISEASE: Defects in RECQL4 are a cause of Rothmund-Thomson
syndrome (RTS) [MIM:268400]. A disease characterized by
dermatological features such as atrophy, pigmentation, and
telangiectasia and frequently accompanied by juvenile cataract,
saddle nose, congenital bone defects, disturbances of hair growth,
and hypogonadism.
CC -1- DISEASE: Defects in RECQL4 are a cause of RAPADILINO syndrome
[MIM:266280]. A disease characterized by radial and patellar
aplasia or hypoplasia.
CC -1- SIMILARITY: Belongs to the helicase family, RecQ subfamily.
CC -1- DATABASE: NAME-Atlas Genet. Cytogenet. Oncol. Haematol.;
WWW="http://www.infobio.fr/services/chromocancer/genes/RECQL4ID285.html".
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC
CC ENBL; AB006532; BAA74453.1; -; mRNA.
DR ENBL; AB026546; BAA86899.1; -; Genomic_DNA.
DR ENBL; BC011602; AAH11602.2; -; mRNA.
DR ENBL; BC013277; AAH13277.2; ALT_INIT; mRNA.
DR HSSP; P15043; 10YY.
DR Ensembl; ENSG00000160957; Homo sapiens.
DR HGNC; HGNC:9949; RECQL4.

DR MIM; 603780; --
DR MIM; 268400; --
DR MIM; 266280; --
DR GO; GO:0003678; F:DNA helicase activity; TAS.
DR GO; GO:0007275; P:development; TAS.
DR GO; GO:0006281; P:DNA repair; TAS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002464; DEAD/DEAH_box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR PANTHER; PTHR113710; RecQ; 1.
DR Pfam; PF00270; DEAD; 1.
DR TIGRFAMs; TIGR00614; RecQ_fam; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; FALSE_NEG.
KW ATP-binding; Catabact; Helicase; Hydrolase; Nuclear protein;
KW Nucleotide-binding; Polymorphism.
FT NP BIND 502 509 ATP (Potential).
FT MOTIF 605 608 DEAD box.
FT VARIANT 267 267 E -> D (common polymorphism).
FT VARIANT 355 355 R -> Q.
FT VARIANT 441 441 P -> S.
FT VARIANT 857 858 Missing.
FT VARIANT 1005 1005 R -> Q (common polymorphism).
FT VARIANT 1208 1208 R -> Q.
SQ SEQUENCE 1208 AA; 133077 MW; CB809A765AB48A1 CRC64;
Query Match 100.0%; Score 6424; DB 1; Length 1208;
Best Local Similarity 100.0%; Pred. No. 2.7e-307;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MERLDVRRERLQAWERAFRRQRPRSQDDVEAAPEETRALYREYRTLKRTTGQAGGLR 60
DB 1 MERLDVRRERLQAWERAFRRQRPRSQDDVEAAPEETRALYREYRTLKRTTGQAGGLR 60
QY 61 SSESLLPAAAEAEPEPCWGPFLNRAATKSPQTPGRSRQGSVPDYQGRKANLKGTLQAG 120
DB 61 SSESLLPAAAEAEPEPCWGPFLNRAATKSPQTPGRSRQGSVPDYQGRKANLKGTLQAG 120
QY 121 PALGRPWPPLGRASSKASTPKPGTGPVPSFAKVSDEPPQLPQPRGRLOHLOASLS 180
DB 121 PALGRPWPPLGRASSKASTPKPGTGPVPSFAKVSDEPPQLPQPRGRLOHLOASLS 180
QY 181 QRLGSLDPGWLQCHSEVPDFLGPACRCPDLGSEESQLLIPGESAVLPGAGSQGPES 240
DB 181 QRLGSLDPGWLQCHSEVPDFLGPACRCPDLGSEESQLLIPGESAVLPGAGSQGPES 240
QY 241 AFOEVSIRVSGPQSSSGGEGKRWNEEPWESPAQVQOESSQAGPPSEGAGAVEEDPPG 300
DB 241 AFOEVSIRVSGPQSSSGGEGKRWNEEPWESPAQVQOESSQAGPPSEGAGAVEEDPPG 300
QY 301 EPVOAOPPCSSPSPNRYHGLSPSSQARAGKAGCTAPLHIFPRLARHGRNRYRLNMKQ 360
DB 301 EPVOAOPPCSSPSPNRYHGLSPSSQARAGKAGCTAPLHIFPRLARHGRNRYRLNMKQ 360
QY 361 KHVVRGRLRSLLRQAWKQKRWKCEGCGGGATVTTKESCFLNEQPDHMAAQCPRA 420
DB 361 KHVVRGRLRSLLRQAWKQKRWKCEGCGGGATVTTKESCFLNEQPDHMAAQCPRA 420
QY 421 SEEDTDAVGPEPLVPSFPVPEVPSLDPTVLPYSLGPSQQLAETPAEVFQALEQLGHA 480
DB 421 SEEDTDAVGPEPLVPSFPVPEVPSLDPTVLPYSLGPSQQLAETPAEVFQALEQLGHA 480
QY 481 FRPQERAVWRILSGISTLLVLTGTAGKSLCYQLPALLYSRRSPCLTLVVSPLSLMDDQ 540
DB 481 FRPQERAVWRILSGISTLLVLTGTAGKSLCYQLPALLYSRRSPCLTLVVSPLSLMDDQ 540
QY 541 VSGLPPLCKAACIHSQMTKQRESVLQKIRAAQVHVHMLTPTPEALVGAGGLPPAAQLPPVA 600

Db 541 VSGLPPCLKAACIISGMRKQBSVLQKIRAAQVHVLMLTPEALVAGGLPPAQLQPPVA 600
Qy 601 FACIDEAHCLSQWHSNFPYCYLRVCKVLRRMGVHCFLGLTATATRTASDVAQHLAAVE 660
Db 601 FACIDEAHCLSQWHSNFPYCYLRVCKVLRRMGVHCFLGLTATATRTASDVAQHLAAVE 660
Qy 661 EPDLHGPAVPTNLHLSVSMRDPTDQALLTLLOKRFONLSIIYCNRRDTERIAALL 720
Db 661 EPDLHGPAVPTNLHLSVSMRDPTDQALLTLLOKRFONLSIIYCNRRDTERIAALL 720
Qy 721 RTCLHAAPVSGGRAPKTTAEAHAGCSEBRRRVQAFMQGQLRVVVAFAFGMLDR 780
Db 721 RTCLHAAPVSGGRAPKTTAEAHAGCSEBRRRVQAFMQGQLRVVVAFAFGMLDR 780
Qy 781 PDVRAVLHGLPPSPESYVQAVGRAGRDGQPAHCHLFLQPGEDLRELRRHVHADSTDFL 840
Db 781 PDVRAVLHGLPPSPESYVQAVGRAGRDGQPAHCHLFLQPGEDLRELRRHVHADSTDFL 840
Qy 841 AVKRLVQRFVFACTCTCTRPSEQEGAVGGERPVKYPQAEQIASHQAAPGPRVCMGH 900
Db 841 AVKRLVQRFVFACTCTCTRPSEQEGAVGGERPVKYPQAEQIASHQAAPGPRVCMGH 900
Qy 901 ERALPIQLTVQALDMPERAIETLLCYLSELPHPHMLLELLATTYTHCRLNCPGPAQLQALA 960
Db 901 ERALPIQLTVQALDMPERAIETLLCYLSELPHPHMLLELLATTYTHCRLNCPGPAQLQALA 960
Qy 961 HRCPPPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDEHPTGV 1020
Db 961 HRCPPPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDEHPTGV 1020
Qy 1021 RRGTVLVEFSELAPHLSPGDLTAEKDKQICDFLYGRVQARERQALRLRTFOAFHSV 1080
Db 1021 RRGTVLVEFSELAPHLSPGDLTAEKDKQICDFLYGRVQARERQALRLRTFOAFHSV 1080
Qy 1081 APPSGCGLQEDDERSTRKDLGRYFEEREGSPGGMEDAQGPQCARLQWEDQVR 1140
Db 1081 APPSGCGLQEDDERSTRKDLGRYFEEREGSPGGMEDAQGPQCARLQWEDQVR 1140
Qy 1141 CDIROFLSLRPEEFSSRAVARIFHGIGSPCYPAQVYQDQRRFWKRYLHLSFHALVGLAT 1200
Db 1141 CDIROFLSLRPEEFSSRAVARIFHGIGSPCYPAQVYQDQRRFWKRYLHLSFHALVGLAT 1200
Qy 1201 BELLQVAR 1208
Db 1201 BELLQVAR 1208

RESULT 2
RECQ4 MOUSE
ID RECQ4 MOUSE STANDARD; PRT: 1216 AA.
AC Q75NR7; Q76MT1; Q99PV9;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE ATP-dependent DNA helicase Q4 (EC 3.6.1.-) (RecQ protein-like 4).
GN Name=Recq14; Synonyms=Recq4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA] (ISOFORM 1) AND NUCLEOTIDE
RP SEQUENCE [MRNA] OF 16-1115 (ISOFORM 2).
RX PubMed=11167012;
RA Ohnata T., Araki R., Fukumura R., Kuroiwa A., Matsuda Y., Tatsumi K.,
RA Abe M.;
RT "Cloning, genomic structure and chromosomal localization of the gene
RL encoding mouse DNA helicase RecQ protein-like 4.";
RN Gene 261:251-258(2000).
RP FUNCTION.

RX PubMed=12915449; DOI=10.1093/hmg/ddg254;
RA Hoki Y., Araki R., Fujimori A., Ohnata T., Koseki H., Fukumura R.,
RA Nakamura M., Takahashi H., Noda Y., Kito S., Abe M.;
RT "Growth retardation and skin abnormalities of the Recql4-deficient
RL mouse.";
RL Hum. Mol. Genet. 12:2293-2299(2003).
RN [3]
RP DEVELOPMENTAL STAGE.
RX PubMed=12952869; DOI=10.1093/hmg/ddg306;
RA Siitonen H.A., Kopra O., Kaeerlaeinen H., Haravuori H., Winter R.M.,
RA Saenaneen A.-M., Peltonen L., Kestila M.;
RT "Molecular defect of RAPADILINO syndrome expands the phenotype
RL spectrum of RECQL diseases.";
RL Hum. Mol. Genet. 12:2837-2844(2003).
RN [4]
RP FUNCTION.
RX PubMed=15703196; DOI=10.1093/hmg/ddi075;
RA Mann M.B., Hodges C.A., Barnes E., Vogel H., Hassold T.J., Luo G.;
RT "Defective sister-chromatid cohesion, aneuploidy and cancer
RL predisposition in a mouse model of type II Rothmund-Thomson
RT syndrome.";
RL Hum. Mol. Genet. 14:813-825(2005).
CC -!- FUNCTION: DNA-dependent ATPase (By similarity). May play a role in
development of the palate and the limbs. May modulate chromosome
segregation.
CC -!- SUBUNIT: Interacts with UBR1 and UBR2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1;
IsoId=Q75NR7-1; Sequence=Displayed;
Name=2;
IsoId=Q75NR7-2; Sequence=VSP_015177;
CC -!- DEVELOPMENTAL STAGE: Not expressed at E12.5. Expressed at E15.5-
E18.5, with highest levels in chondrocytes of developing bone and
cartilage and immature proliferating enterocytes of intestine.
CC -!- MUSCLE-SPECIFIC: Absence of the protein causes early embryonic
lethality. Transgenic mice with exon 13-deleted RECQL4 are severely
growth-retarded and show high (95%) perinatal lethality. They
exhibit various skin, bone, intestine, tooth and thymus
abnormalities and premature aging features, but have normal
sensitivity to IR and UV irradiation. In contrast, transgenic mice
expressing a truncated form of RECQL4 exhibit mild perinatal
lethality, no growth defect, but show defects of the skin and
skeleton, aneuploidy and increased cancer susceptibility.
CC -!- SIMILARITY: Belongs to the helicase family. RecQ subfamily.
CC -!- SIMILARITY: Contains 1 CCHC-type zinc finger.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC -----
DR EMBL; AB039882; BAD11131.1; -; mRNA.
DR EMBL; AB175741; BAD14289.1; -; mRNA.
DR EMBL; AB042529; BAB32696.1; -; Genomic_DNA.
DR HSSP; P15043; 10Y7.
DR Ensemble; ENSMUSG0000033762; Mus musculus.
DR MGI; MGI:1931028; Recq14.
DR PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
DR PROSITE; PS50158; ZF_CCHC; 1.
KW Alternative splicing; ATP-binding; Helicase; Hydrolase; Metal-binding;
Nuclear protein; Nucleotide-binding; Zinc; Zinc-finger.
FT ZN FING 393 410 CCHC-type.
FT NP BIND 519 526 ATP (Potential).
FT MOTIF 627 630 DEAH box.
FT VARSPLIC 843 843
FT
FT CONFLICT 732 739
FT CONFLICT 878 878 S -> G (in Ref. 1; BAD11131).

```
FT CONFLICT 929 929 I -> T (in Ref. 1; BAD11131).
FT CONFLICT 937 937 E -> TK (in Ref. 1; BAD11131).
FT CONFLICT 1111 1111 T -> D (in Ref. 1; BAD11131).
SQ SEQUENCE 1216 AA; 135124 MW; 431D79E43BAEC2AF CRC64;

Query Match
Best Local Similarity 61.4%; Score 3944; DB 1; Length 1216;
Matches 795; Conservative 110; Mismatches 267; Indels 72; Gaps 13;

Qy 1 MERLDVRLQAWERAFRRQRGRPPQDDVEAAPEETRALRYREYRTLKRTTGAGGGLR 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MERLATYRNLQWERAFARLHGERPAKGDVEAAPEETRALRYREYRNLKQVRQADDRH 60

Qy 61 SSE-SLPAARAEAPPCWGHMLNRAATKSPQPTPGRSQGSVDYQORLKANLKGTLQA 119
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 VLEQSLAEAAEEAQEPSCWGPPLSRAATQNTQSPMKQSLSSVDYQKRLKANLKNNTQT 120

Qy 120 GPALGRPWLGRASSKASTPKPGTGPVSPFAEKVSDPEPOLPEQPRGRLOHQAASL 179
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 GPTQSRKLQKQKSLSTVPAPRPPGSKTESPCPDDEADALPRVPEPRPLGOLQQLSSL 180

Qy 180 SQRGLSLDQWLQRCHSEVDFLGAPKACRPDLGSEESQLLIPGESAVLGFAGS----Q 235
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 SRRLLSLDQWLECHNRVSDLLVPGACGLDLSAESQPMQSKVNIAPDPTQSEVSQ 240

Qy 236 GPASAFQEVIRVGSQPQSSSGEKKRMNEEPWESPAQVQOESSQAGPPSEGAGAVVE 295
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 SPEAIAQPAQLVSQSPKINSKGRKKNK-GEDEFAQ-DQPSGAGPLSEGARATVHG 298

Qy 296 EDPGEVQAPQPPQPCSPSPNPRVHGLSPSQAAGAKGTAPLHIIFRLARHDRGNVYR 355
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 QDPPEPTQVNVQPCNS-----SNOARTEKAKGTHLHSPRPAASLDGRGYIR 347

Qy 356 LNMKQHYVRGRLRSRLRLKQAKWKRRKGEFCGGGATVTTKESCFLNEQFDHWAQ 415
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 LNMKQKFFVRVGNRGLLRKQVWKQKQKQAAGSGSPRATDKTCFCRQGFQHWASQ 407

Qy 416 CRRPA-----SEEDTDVAGP-----EPLVSPQVPVPEVPSL 446
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 CSQPGTLTVQEGDRDDKQPISTLEBAVQRTGTASCHSGSEETQPAAPQLQ-VPHCPT- 465

Qy 447 DPTVLPLSLGPSQOLAEPAEVOALEQHQAFRGQERAVNRIISGLISTLLVLTGA 506
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 --PMSPLYPGPLGQVAETPAEVOALERGLYAFRGQERAINRIISGLISTLLVLTGA 523

Qy 507 GKSLCYQLPALLYRRRSPCLTLVSPSLSLMDQVGLPCLKAACITSGWTRKQRESVL 566
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 GKSLCYQLPALLYAQRSPCLTLVSPSLSLMDQVSLDPSCLKAACILHSGWTKQRESVL 583

Qy 567 QKTRAAQVHVLMLTPEALVGA-----GGLPPAAQLPPVAFACIDEAHLCSQWSHNPFCY 621
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
584 KKVRAAQVHVLIIVSPEALVGCARGPGSLPQAAQLPPIAFACIDEVHCLCSQWSHNPFCY 643

Qy 622 LRVCVLRERMGVHCFGLTATARTASDAVQAHLAAVEPDILHGPAPVTNLHLSVMD 681
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 LRVCVLRERMGVHCFGLTATARTASDAVQAHLAAVEPDILHGPAPVTNLHLSVMD 703

Qy 682 RDTQALLTLQGRKRFONLDSIIYCNKREDTERIAALLTCLHAAWVPGSGRAPKTTA 741
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
704 RDSQALVTLQGRDFTLSDVLIYCTRERIONGWLALLTCLISWGDSPRGCGPEAIA 763

Qy 742 EAYHAGCSERRRRVQAFMQGLRVVVAIVAFQMGDLDRPDVRAVLHLGLPPSPESYVQA 801
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
764 EAYHAGMSSQERRRVQAFMRGHLRMVVAIVAFQMGDLDRPDVRAVLHLGLPPSPESYVQA 823

Qy 802 VGRAGRDQPAHCHLFLQPOGEDLRLRRVHADSTDFLAVKRLVQRVFPACTCTCTRP 861
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
824 IGRAGRDGKAHCHLFLMHQPOGEDLWELRRHAHADSTDFLAVKRLVQRVFPCTCS----- 878

Qy 862 SEQEGAVGGERPVPKYPQABQALS-HQAAFGPRRVCWGHERRALPIQLTVQALDMPBEAI 920
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
879 -----QRPVSKSPSEEVKEHSGQOTFVLVQACILGHERALFPVOSTVQALDMEAI 929
```

```
Qy 921 ETLLCYLELHPHWLELLATTYTHCRINCPGGPAQLQALAHRCPPPLAVCLAQQLPEDPGQ 980
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 ETLLCYLELHPHWLELLPWTYAQCHLCLGSSAQLQALAHRCPPPLAACQAKWPPKDTSQ 989

Qy 981 GSSSVEPDMVKLVDSMGWELASVRRALCOLQWDEHPTGVRRGTGVLVFESELAFLHRS 1040
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 GRSSLEFGVVELADSMGWKLASVRQALHQLKWDPEPKGAAGQGTGVLVKFESELAFLHRS 1049

Qy 1041 GDLTASEKQIQCDFLYGRVQARERQALARRRTOAFHSVAFPSGCPCLCQOQDEERSTRL 1100
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1050 GDLTDEEKQIQCDFLYRNVQAREHKAHLHOMSKAFRSVAFPSGCPCLCQOQDEERSTRL 1109

Qy 1101 KDLIGRYFEEEGQEPGGMEDAQGPFGQARLQDWBQVRCDIRQFISLSPKEPKFSRAV 1160
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1110 KTLVSYYFEEEE-EEBETMTDTQPKPGQTQDQWEDQIRRDVRLQLSLRPERFSGRAV 1168

Qy 1161 ARIFPHGIGSPCYPAQVYGODRFRWRKYLHLSFHALVGLATEELL 1204
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1169 ARIFPHGIASPCYPAQVYGLDRFRWRKYLHLDFFHALMHLATEELL 1212

RESULT 3
Q4JNX8_XENLA
ID Q4JNX8_XENLA PRELIMINARY; PRT; 1500 AA.
AC Q4JNX8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE RESQ14-helicase-like protein.
GN Name=RTS;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sangrithi M.N., Bernal J.A., Madine M., Philpott A., Lee J.,
RA Dunphy W.G., Venkitaraman A.R.;
RT "Initiation of DNA replication requires the RESQ14 protein mutated in
RT Rothmund-Thomson syndrome.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; DQ059311; AAY89585.1; -; mRNA.
KW Helicase.
SQ SEQUENCE 1500 AA; 168706 MW; A95F0A2BB28B554D CRC64;
```

```
Query Match 41.4%; Score 2659.5; DB 2; Length 1500;
Best Local Similarity 40.1%; Pred. No. 2.8e-122;
Matches 613; Conservative 206; Mismatches 357; Indels 353; Gaps 35;

Qy 1 MERLDVRLQAWERAFRRQRGRPPQDDVEAAPEETRALRYREYRTLKRT----- 51
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MERYNEVKVLLKWEAFLOEHKEKPKNTDVEASNETQELYKEFKSLKKEQLDSTDV 60

Qy 52 -TQAGGGLRSSSLPAAAEAEPEPR---CWGPHLNRAATKSPQPTGR-SRQGSVPDYG 106
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 FTEKSGQWETSPPKTHRADAGDSEGSASWGTHLNRSVTPKLTSKERDLSKASAQFYG 120

Qy 107 QRLKANLKGTLQAGPALGRR-----PWPLGRASSKASTPKPPG 144
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 MKLKANLGGALREHVPVSLKKSFTRFRPLDSPDVKIKNLNASPKSLDTPSSPLALQDPP 180

Qy 145 TGVVPSFAEKVSDPEPQLPEQPQ---RPGRLQH-LQASLSQRLGSLDPCWLQRC----- 194
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 SEFDTLSHVSLEKPEPHKPMNPPTQGPQIKKQHQLOQAMVGQRMASLNQEWLQRCCLRNG 240

Qy 195 -----HS----- 196
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 VEAMEVTAAEQRTVNRQCNQNTMREALGEGAIEENCKSVKGGPNNDLTHSTQHTVASPKQ 300

Qy 197 -----EVPDF-----LGAPKA 207
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 KVOAYTPVSPHAAAPSLPTTFGRQDTRPENPKLKAEEISDFGRSSSVRSQGNLVSFAI 360
```

```
QY 208 CRPLGSEESQL-----IPESAVLPGAG-----SQPEA----- 239
Db 361 RREBIWPEHQPLRKSKPLRKQICSEK-MSSEAGQVAQLDKSGPMSEHSKRWDQTDSE 419
QY 240 -----SAFORVSI-----RVGSPOPS-----SSGGEKRRNNEEPWE-----SP-AQV 275
Db 420 AMKRTQMGQESLQEVDMQSENIAAEQSVVLAKGKRGKQRETSAPDPTVEVSPKSKK 479
QY 276 QOESSQAGPPSEGAGAV-----AVEEDPPGFPVQAQP---POPCSSPSNPRYHGLSPS 325
Db 480 RRKTNRPDENSESSVRKRKKRGARCMDDSDPAMESPASEQVTKAVSDDLCELEEE 539
QY 326 SQARAG-KAEGTAPLHPPLRLAHDRGNVYVLMKQKHVYGRALRSELLRKAQWKQWR 384
Db 540 PVQSSSVKTVNT-----RMPQIKSENFVRLNKKKSHVKCVKVALNGHRLKQMKQWKQ 592
QY 385 KKGCFGGGGATVT-TKESCFLNEQFDHWAACP----- 417
Db 593 KKGQFGGGGKHFRNSGDTCFRCGGMGHWSQCGSPVVRQEVAEETEEBEVSLPTLED 652
QY 418 -----PASBEDTDAVGEPLVP---SPQVPVEVPSLDPVTPLPLYSGLSPQLAETP 466
Db 653 VARMTNQLRNTNEGLSGWCQDPVTVVNVVAVVEPTPPMPQPLVELGPDGRVKETP 712
QY 467 AEFVQALQHQAPRPGQERAVMRILSGISTLLVLPFGAGKSLCYQLPALLYRRSPCL 526
Db 713 SEVDNLSELGVRSPRQEQAVMRILSGISLSTLSTGMKSLCYQLPAYMYAKRSRCI 772
QY 527 TLVSPPLSLMDQVSGPLCPCKKAACHSGMTRKQRESVLQKIRAAQVHVLMLPEALVG 586
Db 773 SLVISPVLMDQVSGPLCPCKKAVCIHSNNTRTQREAAIEKVKQKRVHVLVLLPEALVG 832
QY 587 AG-----GLPPAAQPPVAFACIDEAHCLSQSHNFRPCYLVRCKVLRMRGMVHCFLGLT 641
Db 833 GGFSSGSLCPADQLPVPVAFACIDEAHCSEWSHNFRCYLRLCKVLRGLGVNCLLGLT 892
QY 642 ATATRTASDAVHAUVAEEDLHGP---APVPTNLHLSVSMRDRTDQALLTLQGRFPQ 698
Db 893 ATATRTAEDVAHLGVSPGEGII--PVRMASVPPNHLHLSVSMRDRDQALLTLKGRFRG 950
QY 699 NLDSTIIICNREDTERIAALRTCLHAANVPGSG-----G 734
Db 951 CLDSVIVVCTREETQRIASLLRTSLQGVTVVSANSHLSQEQEDTTVSDRKAQALKIR 1010
QY 735 RAPKTTABAYHAGMCSRERRRQVAFMGQLRDPVAVVATVAFMGQLRDPVAVVHLGLPSP 794
Db 1011 RPLKWIASEYHAGMSAERRRVQNNFMCGQLLVVATVAFMGQLDKSDVRGIHYNMFKN 1070
QY 795 PESYVOAVGRAGRDGPACHFLFQOQEDLRELRHHVHADSTDFLAVKRLVQRVFPACT 854
Db 1071 PESYVOEIGRAGRDGKDAQCHFLDPEQDMDELARHIYVTSVDYIAVKLVRKVPKCK 1130
QY 855 C-----TCTRPPSEQEGAVGERPVPKYPOE 881
Db 1131 CRDLHREOQALINAGEVDDSELQFLQEGEQDSMCDTPSS-----GRDPVCDTPSG 1183
QY 882 ABOLSHQAAAPRRVCMGHERALPQLTVOALDMPPEAEITLLCYELHPPHWEELLATT 941
Db 1184 AERLCHR-----HERAVPIEETVQSLDIREAEITMCLVLELNPSEHLQLHPT 1232
QY 942 YTHCRNLNCGGPAQIQAALAHRCPPPLAVCLIA-QOLPEDFGQSSSVFEDMVKLVDSMGWEL 1000
Db 1233 LSHCRVVCYSGPQQLRLIARSPPVAVCLARERLAGVDHTHVSTSEFNVELADSMGWEV 1292
QY 1001 ASVRALCOLQWDHPEPTVRGT---GVLVFSELAFLHRSFGDLTAEEKDQICDPLYG 1057
Db 1293 VPKRALDLQWRQPDKGF-KGTGRSGVLVFEFSKLSFHRFSYGBLTDQELDQVCDPLHH 1351
QY 1058 RYQARERQALRLRTFOAFHSAVPSPCGPCLEQOQDEERSTRKDLGLRGYFEEEGQEPG 1117
Db 1352 KVTSRERTSLNQKSCFHAFQSVAVRTSACAERKMSAERSAQKALLLDYFEKRTDTELEA 1411
```

```
QY 1118 GMEDAQPEQOARLOQWEDQVRCIDRQFLSRPERKFSRAVARIFHGIGSPCYPAQVY 1177
Db 1412 KVBDPD--EEYQQIKVQEAQKQVRSRDRHFLSLHIEERFSRALARIFHGIGSPCYPAQVY 1470
QY 1178 GQDRFRWRKYHLHSLFHALVGLATEELIQV 1206
Db 1471 GRDRFRWRKYHLDFEIMNLSKEEIRL 1499

RESULT 4
Q4RLC3 TETNG
ID Q4RLC3 TETNG PRELIMINARY; PRT; 910 AA.
AC Q4RLC3;
DT 13-SEP-2005 (TEMBLrel. 31, Created)
DT 13-SEP-2005 (TEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TEMBLrel. 31, Last annotation update)
DE Chromosome 21 SCAF15022, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0032550001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
NUCLEOTIDE SEQUENCE.
RP Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossaat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castellio V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Bernardis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype."
RL Nature 431:946-957(2004).
[2]
NUCLEOTIDE SEQUENCE.
RP Genoscope; Whitehead Institute Centre for Genome Research;
RG Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
RL -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; CAE01015022; CAG10809.1; -; Genomic_DNA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00487; DEKDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR TIGRfams; TIGR00614; recQ_fam; 1.
DR PROSITE; PS0158; ZF_CCHC; 1.
KW ATP-binding; Helicase; Hydrolase.
FT NON_TER 1 910
SQ SEQUENCE 910 AA; 101865 MW; 7CEFF6939FB4564F CRC64;
```

Query Match 36.0%; Score 2310.5; DB 2; Length 910;
Best Local Similarity 50.9%; Pred. No. 2.3e-105;
Matches 476; Conservative 144; Mismatches 207; Indels 109; Gaps 19;


```
Db 1419 VPGDFTSEIDNALDITLTVSVKQERTQLQLQYVAHGLAAVAYSSCGCCNADFPQDRG 1478
Qy 1098 TRLKDLRLGRYFEEREG-----EPGMEADAQGPQOARLQDWEDOVRCDIRQFLSLRP 1151
Db 1479 EQLKAIVRNFANDYPQDLELEBPSNVPD-----ENIIDVHALINNY 1523
Qy 1152 EEPSSRAVARIFHGISPCYPAQYQODRRFRWKYHLHSFHALVGLATELLQ 1205
Db 1524 DNTFTGRNIARIPIHGIMSPNYPVWGR-CKFWRAHVKVDFNRILHLANNAIIK 1576

RESULT 6
Q9NH11 DROME
ID Q9NH11 DROME PRELIMINARY; PRT; 1530 AA.
AC Q9NH11;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RECQ4.
GN Name=RecQ4; Synonyms=RECQ4; ORFNames=CG7487;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kueano K., Berres M.E., Engels W.R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF233659; AA42939.1; -; Genomic_DNA.
DR HSP; P15043; 10Y.
DR IntAct; Q9NH11; -.
DR FlyBase; FBgn040290; CG7487.
DR GO; GO:000524; P:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; zf-CCHC_1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS0158; ZF_CCHC_1.
SQ SEQUENCE 1530 AA; 169878 MW; 1B753A59A3E30EB4 CRC64;

Query Match 26.2%; Score 1680.5; DB 2; Length 1530;
Best Local Similarity 32.4%; Pred. No. 3.6e-74;
Matches 420; Conservative 191; Mismatches 412; Indels 275; Gaps 36;

Qy 107 ORLKANILKGTLOAGP-ALGRRPWPLGRASSKASTPKPGTGPVPSFAEKVSDRPPQLPEP 165
Db 306 QELQTDNNSSNNQKPDHLNHTP-PASSQKSVAPKNKPP-----PSEQETSDSDSVVAES 359
Qy 166 QPRGRQLQHLQASLSORLSGLDPCWLQ-RCHSEVPDFLGAPKAC-----RPLDGEESQ-- 218
Db 360 EEEGPEYQLSKRKIVSTASKEVVAAPVEIPNKVEPETETTFQENDFDSADEQDA 419
Qy 219 LLTP-----GSAVLPGAGSQ---GPEASAFQEVSVIRGSPQSPSSGGEKKRWNEEPWES 271
Db 420 TYVPENKKDKAKRKQAAGKQKTKPKA-----EPKPKT-----EK 455
Qy 272 PAQVQOESSQAG--PPSEGAGAVAEEDPGEFPVQAQPPQP-----CSSPSNPR 318
```

```
Db 456 KAKVKAEEKPKAEKPRNSKKAIAVEPAPDPEDERQPLNPEDLKVYLAEAGDITSVPR 515
Qy 319 YH-----GLSPSSQARAKAGBTAPLHIFPPLARHD-RGN 353
Db 516 INVQDLEBADATAQRVIRTPAAGPNFGLSEGSNIRVDEKRAAAARKLEERIAAGKLNF 575
Qy 354 VRLNMKQKHYVRG-RALRSRLRLAKQAWKQWR-----KGECEFGGGGATVTTKESC 403
Db 576 VTINIQKKFVRGKSKVNFSSKYKQQRHKKRVAALSGPDMDMGGCDG---VLT---C 628
Qy 404 FLNEQDFHWAQC-----PRP----- 419
Db 629 FQCGGVGHFAQCKVKGDSLLPLSAQLLEDPSFPPTLAEAEQEMASQGVAVHSRNI 688
Qy 420 -----ASEEDTDAVGPEPLVPSQP----- 439
Db 689 QAANAAILQDELNESEEDDOESSEEEVQOHPDPNWSDEMVDVDFALDAVEASLSQP 748
Qy 440 -----VPE-----VPSLDPT-----VLPLYSLGSPGQLAETPA 467
Db 749 VSQEKAAPIKTYVGHKIPEEFLKQAGLDTTASSNSRSHQGGVKPLYDLLPDGVSQDTP 808
Qy 468 EYVQALQEQHQAFRPGQERAVNRILSGISTLLVLTGACKSLCYQLPALLYRRRSCLT 527
Db 809 EVLEALHMFHTNFRKGQDRAIMRTISGLSSVLTSTGSGKSLCYQLPAYLYSRKVGAIT 868
Qy 528 LVVSPILSLMDDQVGLPPCLKAACIHSGMTRKQRESVLQKIRAAOVHVLMLTPEALV-- 585
Db 869 LVISPLVSLMEDQVTVGFHFLRAHCHLTQTAQRMKIQMIANGEIDILLVSPFAVAG 928
Qy 586 ---GAGGLPPAAQLPPVAFACIDEAHCLSQSHGNFRPCYLRCVKLRRMRGVHCFGLT 641
Db 929 ERATGFGAI--LRQLPPIAFACIDEAHCVSQSHGNFRPSYLMICKVLKRLKGLVTLGLT 986
Qy 642 ATATRTASDVAOHLAVB--EPDLHGPAVPVPTNLHLHSVMDRDTQALLTLLOGKRQNL 700
Db 987 ATATLTPRYSIINHGISDGERGIISDIPLPDNLVLSVKDENRDALLQLLNSERFEP 1046
Qy 701 DSIIICNRRDTERIAALLRTCLCHAAWVPGSGRAPKT---TAAAYHAGMCSRERRV 756
Db 1047 QSIIICNRRDTERIAALLRTCLCHAAWVPGSGRAPKT---TAAAYHAGMCSRERRV 1106
Qy 757 QRAFMQQLRVVAVTAVFGMGLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGPACHL 816
Db 1107 QKAFMSNELRIVATIAFGMGINKPDIRAVIHYNMPNPFESYVQIEIGRAGDGLPSHCL 1166
Qy 817 FLOPQGEDLRRLRRVHADSTDFLAVKRLVQRVFPACTCTCTPPSEQEGAVGGERVPVK 876
Db 1167 FLDAKGQDQSELRHRYNSIDRHVIRKLIQKIFVPCSC-----DKEASKRTALP- 1218
Qy 877 YPPQEAELSHQAAPGR-RVCMGHERALPIQLTVQALDMPBEAETLLCYLELPHHWL 935
Db 1219 -----LEGDGRVHMCFGHIGSVKTEVMDLIPAEINISTLLCYMELDPRWC 1267
Qy 936 ELIATTYHCLNCPGPAQLQALAHRCPLAVCLAQQLPEDP--GQSSSVSEFDMVKLVD 994
Db 1268 SVLSSAVMAKVISYGGPKYLKHAACECPPLAMAIQIRDKTFKEDSNIIESVTDIAA 1327
Qy 995 SMGWELASVRRLCOLQWDHEPRTGVRRGTGVLFSELSAFHLRSPGDLTAEKQDQICDF 1054
Db 1328 GIGMNSGVVKYQLKOLEW--VKYNGYPKRSPITVSFVGLGFRICKVPGDFTSEIDNALD 1385
Qy 1055 LYGRVQARERQALRRRTTQAFHSVAFPSGCPGLEQ--DEERSTRKLLILGRYFEEEG 1113
Db 1386 LYTRSVKQERTQLQYVAHGLAAVAYSSCGCCNADFPQDRGEQLKAIVRNFANDYP 1445
Qy 1114 Q-----EPGMEADAQGPQOARLQDWEDQVRCDIRQFLSLRPEEKFSRAVARIFHG 1167
Db 1446 QDLELEBPSNVPD-----ENIIDVHALINNYPDNTFTGRNIARIPIHG 1490
Qy 1168 GSPCYPAQYQODRRFRWKYHLHSFHALVGLATELLQ 1205
Db 1491 MSPNYPVWGR-CKFWRAHVKVDFNRILHLANNAIIK 1527
```

```

RESULT 7
Q7PNM8 ANOQA PRELIMINARY; PRT; 1058 AA.
ID Q7PNM8;
AC Q7PNM8;
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE ENSANGP0000004546 (Fragment)
DE ENSANGP00000000000003572;
GN ORFNames=ENSANGP00000000000003572;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation."
RL Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
preliminary data.
DR EMBL; AA3801008960; EAA11913.3; -; Genomic_DNA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1. CCHC.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; ZF-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00343; Znf_C2HC; 1.
DR TIGRfam; TIGR00614; recQ_fam; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
KW ATP-binding; Helicase; Hydrolase.
FT NON TER 1
SQ SEQUENCE 1058 AA; 117711 MW; 6DFBPF17EA2255FC CRC64;

Query Match 26.0%; Score 1670.5; DB 2; Length 1058;
Best Local Similarity 37.5%; Pred.No. 7.4e-74;
Matches 386; Conservative 147; Mismatches 304; Indels 191; Gaps 24;

QY 321 GLSPSSQARAGKAEPTAPLHIFPRLARHD-RGNYVRLNMKQKHYVRG-RALRSRLLRKQA 378
DB 78 GPSSSGKTKTTVAGSTKETALRKMAAGKLNENFRIDIRKRVFKGKKTINYRYKSKQ 137
QY 379 WKQKRVKGEFCFG-----GGATVTTTKESCFLNEQPDHWAACPRPASEE---DTD 426
DB 138 WK---AKKAAALTGPMDNRGCDGGTLT---CFQCGGTGHMAFCKPLEDKLLPYDAD 190
QY 427 AVGPE-----PLVPSP-----QPVE 442
DB 191 TIESSFPPTLEBAEAMANSRTLVHNSRIERLPVANPVWKEQEQBDLPABEEVGEPGKE 250
QY 443 VPSLDP----- 448
DB 251 EPSTLPVEDDLDEPMEQSDHQTEQNAASQAPAPAYIGHKIPPEFLKQSGLLVTV 310

```

```

QY 449 -----TVLPYLSLPGSQLAETPAEVFOALEQLGHOAFRPGOBERAYVMRIILSGISTLVL 502
DB 311 SGGFKGTVEPLYQORGDCTVPTTPEVFEALRMFGHOSFRHQGERAVNRVLCGLUSTLVL 370
QY 503 PTGAGKSLCYQLPALLYRRSPCLTLVVSPLLSLMDQVSLGPLPCLKAACIHSQMTKRQR 562
DB 371 STGAGKSLCYQLPAYLYRKNRSCITLVIPLVSLMEDQVHGMPDFLNAHCLHTTQTPKVR 430
QY 563 ESVLQKTRAAQVHVIMLTPEALV-----GAGGLPPAAQLPPVAFACIDEAHCLISQSHN 616
DB 431 DRTMQAIAAGEVDVLLISPEAVVSGEKSTGFGSL--LRQLPPIAFACIDEAHCVSQSHN 488
QY 617 FRPCLRVCKVLRRMGVHCFGLGTATATRTASDVQAHLAVAEPPD---LHGPAVPPT 672
DB 489 FRPSYLMICKVLEKKGKVTILGTATATVQTRQISVSHLAI---PDGVQGIISDIPDP 545
QY 673 NLHLSVSMRDRTDQALLTLQGRFQNLDSIIYYCNRRREDTERIAALLRTCHLAAPVGS 732
DB 546 NLLLTVSERDANRDVALVELLQSERFSLSIQSIIVYCTRRDDCERVATFTRTCQDAARAA 605
QY 733 GGRAPKT-----TAEAYHAGCRRRRRVRQAFMQGLRVVAVAFQMGMLDRDPDRAVL 787
DB 606 AANAHKRKLNYVAEPYHAGMPASRRRTIQNAFMMSGELRIVVATIAFGMGINKADIRAI 665
QY 788 HLGLPSPFESVQAVGRAGRDGQPAHCHLFLQPOGEDLRELRHHVHADSTDFLAVKRLVQ 847
DB 666 HYNPKNFESVQVGRAGRDGLLHCHLFLDGKGRDNRNLRRIYANSIDRHVIRKLLQ 725
QY 848 RVFPACTCTCTRPPEQ-----EGAVGGERPVPKYPQEAQSLHQAAQPPRRVCMGHE 901
DB 726 KIFVPCACA---KSHQOVILFTDGGGGGT-----KRLCPGHE 760
QY 902 RALPIQITVQALDMPERAEITLLCYLELHPHHWLELLATTYTHCLNCPGPGPAQLQALAH 961
DB 761 ICFSEIATVQQLDIPENITTFCLYLELDEQRYIQAQSPAYTMCKVMSYGVVPLRQAQ 820
QY 962 RCPPLAVCLAQLPEDPGQ--SSSVFEDMVKLVDMSGMWELASVRRALCOLQMDHEPRGTG 1020
DB 821 ECPFLAMAFALDLKRGISHATSTATIEFPVIDVASAIGMDSGVVKQLKNLEW---TTVNV 878
QY 1021 RRGTVLVEFSELAFLHRS PGDLTAEEKDQICDFLYGRVQARERQALRLRRTFOAFHSV 1080
DB 879 RKRSPLSVTFTFELGFRVAPGDLTDELDHALDGLYERVTHQERTQLAQLQYISDALNSV 938
QY 1081 APPSCGPLEQOD--HEKSTRLKDLLGRYFEEEGQEPGGMEDAQGPQGQARLODWDQ 1138
DB 939 CFNTIGP-VSRADCPGSDKLTIVREYFTTDSKE---QIEIVPEPDDTT----DEQ 989
QY 1139 VRCDIRQFLSLRPEKPSRAVARIFHGIGSPCYPAQVYQDRRFRWKYHLHSHALVGL 1198
DB 990 LINDIRTAICRYPENFTGTRAIARLPHGVQSPNYTALVMSRS-NFWRAYTKTDFNRI 1048
QY 1199 ATEELLQV 1206
DB 1049 ANAEIVRM 1056

RESULT 8
Q7YR85 BOVIN PRELIMINARY; PRT; 340 AA.
ID Q7YR85_BOVIN;
AC Q7YR85;
DT 01-OCT-2003 (TremBLrel. 25, Created)
DT 01-OCT-2003 (TremBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE RecQ protein-like 4 (Fragment).
GN Name=RECQL4;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]

```

```
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14667821; DOI=10.1016/S0888-7543(03)00238-6;
RA Winter A., Alzinger A., Fries R.;
RT "Assessment of the gene content of the chromosomal regions flanking
RL bovine DCAT1.";
RL Genomics 83:172-180(2004).
DR EMBL; AJ518973; CAD58810.1; -; Genomic_DNA.
DR HSP; P15043; 10YV.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00490; HELIC; 1.
FT NON TER 1
FT NON TER 340
FT NON TER 340
SQ SEQUENCE 340 AA; 37366 MW; 205F057DD109CAED CRC64;

Query Match 18.2%; Score 1172; DB 2; Length 340;
Best Local Similarity 67.8%; Pred. No. 6.6e-50;
Matches 232; Conservative 27; Mismatches 57; Indels 26; Gaps 2;

Qy 601 PACIDEAHCLSQSHNRPFCVYCKV-----LRRMGVHC 636
Db 1 FACIDEAHCLSQSHNRPFCVYCKVSLGGGSGVQREGTAGSPPPSOTLRDQMGVHC 60

Qy 637 FLGTATATRTASDAQAHLAVAEEDPLHGPAPVPTNHLHLSVSDRDTDOALLTLQGR 696
Db 61 FLGTATATRTASDAQAHLAVAEEDPLHGPAPVPTNHLHLSVSDRDTDOALLTLQGR 120

Qy 697 FQNLDSIIICNRREDTERIAALRTCLHAAVPGSGRAPKTTAAEYHAGMCSRERRV 756
Db 121 FRALGSVIIICNRREDTERIAALRTCLHAAVPGSGRAPKTTAAEYHAGMCSRERRV 180

Qy 757 QRAPMOQLRVVATVAFMGMLDRPDVRAVLHGLPSPFESYVQAVGRAGRDGQPAHCHL 816
Db 181 QRAPMEGRRLMVVATVAFMGMLDRPDVRAVLHGLPSPFESYVQAVGRAGRDGQPAHCHL 240

Qy 817 FLOPQGDRLRLRHVADSTDFLAVKRLVQRVEPACTCTCTPPSQEAGVGRVVPK 876
Db 241 FLRPOGDLRRHRHVANVDFFAVRLVQRVEP--PCACARQPPQEGSRSEEGHLAG 298

Qy 877 YPOEAEQLSHQAAPGPRRCVCMGHERALPIQLTVQALDMPPEE 918
Db 299 APVAASAQDSQSPVHTPCPGHERVLPVQTVQALDMPPEE 340

RESULT 9
Q7X829 ORYSA
ID Q7X829 ORYSA PRELIMINARY; PRT; 927 AA.
AC Q7X829;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE QJ991113 30.2 protein.
GN Name=QJ991113 30.2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=2447439; PubMed=12447439; DOI=10.1038/nature01183;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Weng Q., Zhang L., Lu Y., Mu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Yu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
```

```
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
CC -1- FUNCTION: EIF4A is both a subunit of a high molecular weight
CC protein complex involved in cap recognition and is required as a
CC single polypeptide chain for mRNA binding to ribosome. It is an
CC ATP-dependent single stranded DNA-binding protein with a sequence-
CC independent unwinding activity (helicase) (By similarity).
CC -1- SUBUNIT: eIF4P is a multi-subunit complex, the composition of
CC which varies with external and internal environmental conditions.
CC It is composed of at least EIF4A, EIF4G, EIF4S and EIF4G (By similarity).
CC EMBL; AL662946; CAD41320.2; -; Genomic_DNA.
DR HSP; P15043; 10YV.
DR Gramene; O7X829; -.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR ATP-binding; Helicase; Hydrolase; Initiation factor;
DR Protein biosynthesis.
DR SEQUENCE 927 AA; 102031 MW; 48B7BC5504C04FFA CRC64;

Query Match 15.1%; Score 971.5; DB 2; Length 927;
Best Local Similarity 26.5%; Pred. No. 1.4e-39;
Matches 303; Conservative 168; Mismatches 392; Indels 279; Gaps 37;

Qy 110 KANLKGTLQAGPALGRRPWPGLGRASSKASTPKPP---GTGVPSPFAKVDSEPPQLPEPQ 166
Db 13 EGSLLSDVSASPPRRRSP---PRPAPPPPPPPPKHTRTPAAPTKPKLPTPPAASAPA 69

Qy 167 PRGRLLQHLQASLSQRLGSLDP--GWLQRCHSEVPDFLGAPKACRPDLGSEES---QLLIP 222
Db 70 PAPPPPTLRAAALS-----DPHGLAAR-----IAAGSALTAAGTASSSSFRLLVQ 116

Qy 223 GESAVLPGAGSQGPASAFQEVYSIRVSGSPQSSSGGKKRRMNEEPWESPAQVQESSQA 282
Db 117 SRNPSFDPATFTAPASSAPSEVPSAAPRPPTAA-----TDAFPQTRPKRVHP 165

Qy 283 GPPSEGAGAVAEEDPPGEPVQAPQPPQPCSPSPNRYHGLSPSSQAPAGKAETAPLHIF 342
Db 166 NSVSEVAAAASAAAEQP-----KRAAGG----- 187

Qy 343 PRLARHDRGNVRLNWKQKHVYGRALRSLRLKQAKWKKGECFGGGGATVTTKES 402
Db 188 -----SEGNFVRLNI--NGYGRRTFKNSQAKRSTKCRSWRQ-----RAAGAT----- 229

Qy 403 CFLNEQFDHWAACQPRPASEEDTAVGPPEPLVPSPQVPVPSLDPTPLVLSLPSGSQL 462
Db 230 -----PRSQGDEGDLVAELLEREKQAASD-----SVLEAVE-----SV 264

Qy 463 AETPAEVFOALEQL-----GHQAFRFGQERAVNRILSGISTLLVLPFGAGKSLCY-OLPA 516
Db 265 REDPSE--QNLKSLNAAAYGHDSFRQGLEAIQQIVAGESTMLVLPFGAGKSLCYQCVPA 322

Qy 517 LLYRSRSPCLTVVSPLLSIAMDQVSLPCLAKAACIHSQWTRKQRESVLQKRAAOVHV 576
Db 323 MIL-----PGLTVVSPLLSIAMDQVSLPCLAKAACIHSQWTRKQRESVLQKRAAOVHV 378

Qy 577 LMLTPEALVGAGGLPPAAQLPPVAFACIDEAHCLSQSHNRPFCVYCKV-----LRRMGVHC 635
```

Db 379 LFVSPERFLNEEFLLIFRDTLPISLVAIDEAHCISEWSHFRPSYLRASLRKLNQV 438
 Qy 636 CFLGLTATARTASDVAAQLAVAEEDLHGPAPVPTNLHLSWMDRDTQALLTLQK 695
 Db 439 CILAWTATATQTLLEIMWALEIFSD-NLIQTSQIRENLQLSISTDNRLKDLMLLKSP 497
 Qy 696 RFQNLDSIIICNRRDTERIAALLTCLHAAWVPGSGRPAKTTAAAYHAGMCSRERR 755
 Db 498 PFVDMRSIIIVYCKFAETDFVSKYL--CDN-----NITAKSYHSLGILKRSR 543
 Qy 756 VQRAFQMLQVAVVATVAFGMLDRPDVRAVLHLGLPPSPESYVQAVGRAGRDGPAHCH 815
 Db 544 VOELFCNKIPQBAEQSHQAAPGPRVCMGHERALPIQLTVQALDMPHAEITLCLYLH 930
 Qy 816 LFLOQGEEDLRELRHHADSTFLAVKRLVQRYFPA-----CTCTCTRPPEQEGAVG 870
 Db 604 LLL--DSATFYKIRSLSHSDGVGAMSKFLYQIFSSENTTGCICSLAK----- 650
 Qy 871 ERPVKYPPQBAEQSHQAAPGPRVCMGHERALPIQLTVQALDMPHAEITLCLYLH 930
 Db 651 VOELFCNKIPQBAEQSHQAAPGPRVCMGHERALPIQLTVQALDMPHAEITLCLYLH 930
 Qy 931 PHWLELLATTYTHCRNLCPGPAQLALAHRCPLAVC-----LAQQLPEDPGQSSSVE 986
 Db 675 DQYIRLLPQPSVTCCTL-----YFHTSPQLLADKDIILRSVLRSEMKDGHYV 723
 Qy 987 FDMVKLVDSMGWELASVRRALCOLQWDEHPTGVRRTGTVGVFVS-ELAFHLRSPG--- 1041
 Db 724 FDIPIRIANDLKITWNEV-----FDHLK-----LKFSGEISFELKDPAYCYV 765
 Qy 1042 -----DLTAEKDOI CDLYGRVQARERQALRLRTTFQAFHSAVPSC----- 1085
 Db 766 ILWRPDDFNALSAN-----LTKWLSVESSKISKLDAMF-ALANFAVKGCKRTGCGSGQ 819
 Qy 1086 -GPCLEQDEERSTRLLDGLGRYFEEBEGQEPGGMEDAQGPQARLDQWEDOVRCDIR 1144
 Db 820 HTPCIQKIME-----YFSKDDGTSENC-----RTQLQKSSPFLQADIK 859
 Qy 1145 QFLSLRPREKFSRAVARIFHGIGSPCYPAQVYGQDRFRFKYHLHLSHALVGLATELL 1204
 Db 860 VPIQSNFAKTPTRAVARIMHGISSPAPSVTWSKN-HFWGRYVEVDPLVWEAKAEVL 918
 Qy 1205 QV 1206
 Db 919 KL 920

RESULT 10

Q8W028_ARATH PRELIMINARY; PRT; 870 AA.
 AC Q8W028;
 DT 01-MAR-2002 (TREMELrel. 20, Created)
 DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE Helicase.
 GN Name=rdl-5;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RA Hartung F., Pichova H., Puchta H.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: eIF4a is both a subunit of a high molecular weight
 CC protein complex involved in cap recognition and is required as a
 CC single polypeptide chain for mRNA binding to ribosome. It is an
 CC ATP-dependent single stranded DNA-binding protein with a sequence-
 CC independent unwinding activity (helicase) (By similarity).
 CC -!- SUBUNIT: eIF4F is a multi-subunit complex, the composition of
 CC which varies with external and internal environmental conditions.

CC It is composed of at least eIF4A, eIF4E and eIF4G (By similarity).
 DR EMBL; AJ421618; CAD13472.1; -, mRNA.
 DR HSP; P15043; 10Y.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0003743; F:translation initiation factor activity; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR002464; DEAH_box_C.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004589; RecQ.
 DR Pfam; PF00271; DEAD; 1.
 DR Pfam; PF00271; Helicase_C; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR TIGRFAMs; TIGR00614; recQ_fam; 1.
 DR PROSITE; PS00690; DEAD ATP HELICASE; UNKNOWN_1.
 KW ATP-binding; Helicase; Hydrolase; Initiation factor; Nuclear protein;
 KW Protein biosynthesis.
 SQ SEQUENCE 870 AA; 96383 MW; 8B5ADA4F4A81723 CRC64;

Query Match 14.0%; Score 897.5; DB 2; Length 870;
 Best Local Similarity 26.7%; Pred. No. 5.6e-36;
 Matches 295; Conservative 160; Mismatches 368; Indels 281; Gaps 38;

Qy 133 ASSKASTPKPGTGPVPSFAEKVSDPEPPQLPEQPRGRQLQHLQASLSQRLSLDPLQ 192
 Db 3 SSSRSRKPAP-THPPNPSPQEAAPVPSPPPPSP-----LFTNLPFRICQSPA--- 53
 Qy 193 RCISEVDFLGPACACAPDLGSEESQLLPGESAVLPGAGSQGPSEAFSEVIRVGS 252
 Db 54 RFSSSVSPF---SRLCSRASFTSVKLSGDGVFV-----PEPLVEVIA----- 95
 Qy 253 QPSSSGGKRWNEEPWESPAQVQSSQAGPSEGAGAVEDDPG---EPVQAQPPQ 309
 Db 96 -----PPK-----SVRRKPNLITDITSPVK 118
 Qy 310 PCSSPSNPRVHGLSPSSQARAGKAGTAPLHIFRLARHGRGNVLRNMKKHYVRGAL 369
 Db 119 PMWFRN-----GNCE-----GNFVKLNNGK---RGKFF 145
 Qy 370 RSRL-----LRQAWKQKWKKEGCGGGATVTTKESCFNMQFDHWAACQPRPASEE 423
 Db 146 PSKYGVSKSRSSYFGRKYEADGDGSL--EESDLQKI-----ED 191
 Qy 424 DTDVAGPEPLVSPQVPEVPSLDPTVLFLYSLGSPGQLAETPAEVFOALQGHQAFRP 483
 Db 192 EANGF-----ISSVEDAILAVKTEASDENLTLLNLVY-----GYDSFRD 231
 Qy 484 GQERAVNRILSGISTILVLPFGACKSLCYQLPALLYRSRSPCLTYVSPILSLMDQVSG 543
 Db 232 GLOQAKIMILGSGSTMLVLPFGACKSLCYQIPAMIL-----PGITLWVSPILSLMDQ 287
 Qy 544 LPPCLKAACITSGMTRKQRESVLQIRAAQVHVLMLTPEALVGAGGLPPAAQLPVPVAFAC 603
 Db 288 LPSIKKGLSSSSQRPPEATEETLRKKEGIIKVLVSPERLLNVFELSMFMSLSVLV 347
 Qy 604 IDEAHCLSQWNSHNPVPCYLRV-CKVLRMRMGVHCFGLGTATATRTTASDVAAQLHVA 662
 Db 348 VDEAHCVSEWSHNPVPSYMLKASMLFSELKACILAMTATATMTLQAVMSLEI-PST 406
 Qy 663 DLHGAPVPTNLHLVSM---DRDTQALLTLQKRPONLDSIIICNRRDTERIAL 719
 Db 407 NLIQKSLQDNFELSVLSGANRMD--LLTIMESPPYKEIRSIIVYCKFYQETDMISKY 464
 Qy 720 LRTCLHAAWVPGSGRAPKTTAAEYHAGMCSRERRRVRQAFMQGLRVVAVTAFGMGLD 779
 Db 465 LRD-----NNINAKGVHSLGPKADRVRIQESFCSNKIRVVAVTAFGMGLD 510

QY 901 ERALPIQUTVQALDMPERAIETLTCYLELHPHHLELLIATTTTCRLNCPGPPAQLQALA 960
 Db 651 --SLVIESASQKFMKEVMQITLTHLGEVQVLRMLPQLNICTLN-----F 697
 QY 961 HR--CPLPLAVCLAQQLPEDPGQSSSVPEPDMVKLVDSMG---WELASVRRALCOLQWDH 1014
 Db 698 HKTSDAQINVMLYVQSSNTLAARSAIVAAILKSHVKQGLHVFDPAPVASSICVAT--- 754
 QY 1015 EPRTGVRGTCGLVFEFSELAFLHRSFGDLTAEEKD-----QICDFLYGRVQAR 1062
 Db 755 -----TDVLAIEIQAL-----KGEVTVELKDSAFCTILKSPKXICSLSHLTKWL 799
 QY 1063 ERQALARLR--TFQAFHVSVPSCPCLEQQDEERSFLKDLLGRYPPEEESQSPGGMED 1121
 Db 800 TEIESCKVRKLDIMSSAAVAISVNTSELSSGAKQTR--SLQGRIFDYFNG-----D 850
 QY 1122 AQGPBGQARLQWEDQ-----VRCDIROFLSLRPEEKFSRAVARIFHGIGSPCVPAQVY 1177
 Db 851 EKCSUSPSKA-----TQCAFLRADIKVFLQSNROAKFTTPRAIAKIMGVGSPAPFNSV 904
 QY 1178 GQDRRFWRKYLHLSPHALVGLATEL 1203
 Db 905 SK-THFWGRYVMVDFRIMEAAQTLEL 929

RESULT 12
 Q7FAH0 ORYSA
 ID Q7FAH0 ORYSA PRELIMINARY; PRT; 874 AA.
 AC Q7FAH0
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE QJ000223 09.17 protein.
 GN Name=QJ000223 09.17;
 OS Oryza sativa [japonica cultivar-group].
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriartoideae; Oryzeae; Oryza.
 OC NCBI Taxid=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=23337377; PubMed=12447439; DOI=10.1038/nature01183;
 RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
 RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Yang K., Yu S., Tang Y.,
 RA Weng Q., Zhang L., Lu Y., Wu J., Lu Y., Zhang L.S., Yu Z., Fan D.,
 RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
 RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
 RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
 RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
 RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
 RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
 RA Han B.;
 RT "Sequence and analysis of rice chromosome 4.";
 RL Nature 420:316-320(2002).
 CC -!- FUNCTION: E1F4A is both a subunit of a high molecular weight
 CC protein complex involved in cap recognition and is required as a
 CC single polypeptide chain for mRNA binding to ribosome. It is an
 CC ATP-dependent single stranded DNA-binding protein with a sequence-
 CC independent unwinding activity (helicase) (By similarity).
 CC -!- SUBUNIT: e1f4p is a multi-subunit complex, the composition of
 CC which varies with external and internal environmental conditions.
 CC It is composed of at least E1F4A, E1F4E and E1F4G (By similarity).
 CC EMBL; AL606998; CAD41405.2; -; Genomic_DNA.
 DR Graneine; Q7FAH0; -;
 DR GO; GO:0005524; F-ATP binding; IEA.
 DR GO; GO:0008026; F-ATP-dependent helicase activity; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0003743; F:translation initiation factor activity; IEA.
 DR GO; GO:0006259; P:DNA metabolism; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR InterPro; IPR011545; DEAD/DEAH_N.
 DR InterPro; IPR004589; RecQ.

DR Pfam; PF00270; DEAD; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR TIGRFAMs; TIGR00614; recQ_fam; 1.
 KW Helicase; Initiation factor; Protein biosynthesis.
 SQ SEQUENCE 874 AA; 96275 MW; 51B64C486617774F CRC64;
 Query Match 11.8%; Score 761; DB 2; Length 874;
 Best Local Similarity 23.8%; Pred. No. 2.9e-29;
 Matches 271; Conservative 163; Mismatches 377; Indels 330; Gaps 38;
 QY 110 KANLKGTLOAGPALGRPPWPLGRASSKASTKPP---GTGPVPSFAEKVSDPEPQLPEQP 166
 Db 13 EGSLLSDVSASPPRRSP---PRPAPPPPPPPPPGHRTPAAPATPKLKPPTPPAASAPA 69
 QY 167 PRPRQLHLOASLSQRLGSLDP-GWLORCHSEVPDFLGAPKACRPDLGSEES---QLLIP 222
 Db 70 PAPPPTPLRAAALS-----DPHGLAAR-----IAAGSALTAAGTASSSSFRRLVQ 116
 QY 223 GESAVLFGAGSQGPASAFQEVSIKRVGSPSSSGCKRRWNEEPWESPAQVQOESSQA 282
 Db 117 SRNPSFDPAFTAPASAPSEVPSAAPPPTAA-----TDAPPQTPKRVHP 165
 QY 283 GPPSEGAGAVAEEDPPEEPVQAQPPQPCSSPNRYHGLSPSSQARAGKAETAPLHIF 342
 Db 166 NSVSEVAAAAGAAEQP-----KRAAGG----- 187
 QY 343 PRLARHGRNYVRLMKQKHYVRGRALRSLLRKQAWKQKWKKEGCEGCGGATVTTKES 402
 Db 188 -----SEGNFVRLNI--NGYGRRTTFKNSQAKSTKCRSWRRQ-----RAAGAT- 229
 QY 403 CFLNEQDHWAAQCPRPASBEDTDAVGPELVPSPQVPPEVPSLDPTPLVLSLPGSQ 462
 Db 230 -----PRSGDEGLVAEALLERKQAASD-----SVLEAVE-----SV 264
 QY 463 AETPAEVFQALEQL-----GHQAFRPGQERAVMRILSGISTLLVLPTGAGKSLCYQLPAL 517
 Db 265 REDPSE--QNLKSLNAAAGHDSFRQGLEAIQIIVAGESTMLVLPTGAGKSLCYQVDM 322
 QY 518 LYSRRSPCLTLVSPFLSLMDQVSGPLPPCLKAACIIHSGMTRKORESVLQIRAAQVHL 577
 Db 323 IL-----PGLTLVSPFLSLMVDLRLKLPFAFPGGLLASSQTSDEFHDTLQRLAGEIKVL 378
 QY 578 MLTPEALVGGLPPAAQLPVPVAFACIDEAHLCSQWSHNPFRPCYLVR--CKVLRERMGVHC 636
 Db 379 FVSPERFNEEFLLIFRDTLPILVAIDEAHCISEWNSHNPFRPSYLRRLASLRRLKLVQC 438
 QY 637 FLGTATATRTASDVQAQLAVABEPDLHGPAVPVPTNLHLSVSMRDTDOALLTLQGRK 696
 Db 439 ILAMTATATTQTLBEIMNNALEIPSD-NLIQTSQIRENLIQLSISTSDNRLKDLMLLKSP 497
 QY 697 FQNLDSIIYCNRRREDTERIAALLRCTLHAAWVPGSGRPAKTTAAAYHAGMCSRERRV 756
 Db 498 FVDMRSIIYCKFOAETDFVSKYL--CDN-----NITAKSYHSGLLIKNRGRV 543
 QY 757 QRAFWQGLRVVTVATVAFMGMLDRPDVRAVLHLGLPSPSFVQAVGRAGRDGPAHCHL 816
 Db 544 QELFCSNKRIVF-----ST-----NF 559
 QY 817 FLQPGQEDLRLRRHVHADSTDFLAVKRLVORVFA-----CTCTCTPPSPSEGEAGVGE 871
 Db 560 TEQP-----FLGSDGVDGYAMSKFLQIFSENTTGCISLAK----- 597
 QY 872 RPVPKYPQEAQLSHQAAPGPRVCMGHERALPQLTVQALDMPERAIETLLCYLELHP 931
 Db 598 -----ELTSRKFDIKEEVLLTILTQLEICD 622
 QY 932 HHWLELLATTYTHCELNCPGGAQIQAALAHRCPPPLAVC-----LAQLPEDPQCGSSSVF 987
 Db 623 QOYIKLLPQFSVTCTL-----YFHKTSIPOLLADKDLILRSVLRNLRSEMKDGHVYVF 671
 QY 988 DMVKLVDSMGWELASVRRALCOLQWHDHPRGTGVRRTGVLGVVEFS-ELAFHLRSPG----- 1041
 Db 672 DIPRIANDLKITMNEV-----FDHLHK-----LKFSGEISFELKDPAYCYVI 713

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA], AND SEQUENCE REVISION
 TO 254.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503; DOI=10.1126/science.277.5331.1453;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RN Science 277:1453-1474 (1997).
 [4]
 RP SEQUENCE REVISION TO 502-503.
 RA Arnaud M., Berlyn M.K.B., Blattner F.R., Galperin M.Y., Glasner J.D.,
 RA Horiuchi T., Kosuge T., Mori H., Perna N.T., Plunkett G. III,
 RA Riley M., Rudd K.E., Serres M.H., Thomas G.H., Wanner B.L.;
 RT "Workshop on annotation of *Escherichia coli* K-12.";
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 [5]
 RP FUNCTION, AND PROTEIN SEQUENCE OF 1-5.
 RX MEDLINE=90311913; PubMed=2164680;
 RA Umezaki K., Nakayama K., Nakayama H.;
 RT "Escherichia coli RecQ protein is a DNA helicase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5363-5367 (1990).
 CC -!- FUNCTION: Involved in the recF recombination pathway; its gene
 CC expression is under the regulation of the SOS system. It is a DNA
 CC helicase.
 CC -!- INTERACTION:
 CC P06958: aceE; NbExp=1; IntAct=EBI-552058, EBI-542683;
 CC P06959: aceF; NbExp=1; IntAct=EBI-552058, EBI-542707;
 CC P28905: holC; NbExp=1; IntAct=EBI-552058, EBI-549159;
 CC P28632: holD; NbExp=1; IntAct=EBI-552058, EBI-549176;
 CC P00391: lpdA; NbExp=1; IntAct=EBI-552058, EBI-542856;
 CC P0A7Y4: rnaA; NbExp=1; IntAct=EBI-552058, EBI-556910;
 CC P02370: rpsN; NbExp=1; IntAct=EBI-552058, EBI-543971;
 CC P28043: sbsB; NbExp=1; IntAct=EBI-552058, EBI-549252;
 CC P06612: topA; NbExp=1; IntAct=EBI-552058, EBI-544172;
 CC P14294: topB; NbExp=1; IntAct=EBI-552058, EBI-552080;
 CC -!- SIMILARITY: Belongs to the helicase family. RecQ subfamily.
 CC -!- SIMILARITY: Contains 1 HRDC domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; M30198; AAA24517.1; ALT_INIT; Genomic_DNA.
 CC EMBL; M87049; AAA67618.1; ALT_INIT; Genomic_DNA.
 CC EMBL; U00096; AAT48221.1; -; Genomic_DNA.
 CC PDB; 1OYW; X-ray; A=1-522.
 CC PDB; 1OYV; X-ray; A=1-522.
 CC IntAct; P15043; -;
 CC EcoBASE; EB0826; -;
 CC EcoGene; EG10833; recQ.
 CC InterPro; IPR006293; ATP RecQ.
 CC InterPro; IPR001410; DEAD.
 CC InterPro; IPR011545; DEAD/DEAH_N.
 CC InterPro; IPR002464; DEAD/DEAH_N.
 CC InterPro; IPR001650; Helicase_C.
 CC InterPro; IPR002121; HRDC.
 CC InterPro; IPR004589; RecQ.
 CC PANTHER; PTHR13710; RecQ; 1.
 CC Pfam; PF00270; DEAD; 1.
 CC Pfam; PF00271; Helicase_C; 1.
 CC Pfam; PF00570; HRDC; 1.
 CC SMART; SM00487; DEXDC; 1.
 CC SMART; SM00490; HELIC; 1.
 CC SMART; SM00341; HRDC; 1.
 CC TIGRFAMs; TIGR01389; recQ; 1.
 CC TIGRFAMs; TIGR00614; recQ_fam; 1.
 CC PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NBG.
 CC PROSITE; PS00967; HRDC; 1.

KW 3D-structure; ATP-binding; Complete proteome;
 KW Direct protein sequencing; DNA damage; DNA recombination; DNA repair;
 KW DNA-binding; Helicase; Hydrolase; Nucleotide-binding; SOS response.
 FT INIT MET 0
 FT DOMAIN 528 608 HRDC.
 FT NP BIND 46 53 ATP (Potential).
 FT MOTIF 145 148 DEAH box.
 FT CONFLICT 254 254 R -> A (in Ref. 1 and 2).
 FT CONFLICT 502 503 RG -> A (in Ref. 1 and 2).
 FT CONFLICT 599 599 R -> P (in Ref. 1).
 SQ SEQUENCE 608 AA; 68232 MW; AA01654270061AEI CRC64;
 Query Match 9.8%; Score 631.5; DB 1; Length 608;
 Best Local Similarity 37.6%; Pred. No. 4.4e-23;
 Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;
 Qy 474 EQLGHQAFRPGQERAVNRILSGISTLLVLP TGAGKSLCYQLPALLYSRSPCLFLVVSPL 533
 Db 18 ETFGYQQFRPQEEIIDTVLSGRDCLVVMPTGGKSLCYQIPALLNG---LTVVVSPL 73
 Qy 534 LSLMDDQVSGLPP-CLKAACIHSGMTKQRESVLQKTRAAQVHVLMLTPEALVGAGGLPP 592
 Db 74 ISLMKDQVDQLQANGVAAACINSTQTRQQLVMTGCTGQIRLLYIAPELMLDNFLEH 133
 Qy 593 AAQLPPVAFACIDEAHCLISQSHNFRPCYLRCVKLVLRMGVHCFLGTATATRTASDV 652
 Db 134 LAHNPVLLA-VDEAHCISQGHDFRPEYAAALGQ-LRQRPFTLPFMALTATADTTQDI 191
 Qy 653 AQHLAVAEPLDHPAPVPTNHLHLSVMSDRDTPDQALLTLQGRKFNLDLSI----- 703
 Db 192 VRLLG-----LNDPL-----IQIS-SFDRPNRYMLM---EKFKPLDQLMRVYVEQGR 235
 Qy 704 ---IYCNRRDTERIAALLTCLHAAVPGSGGRAPKPTTAEYAHGMSRRRVRQAF 760
 Db 236 KSGIYCNRAKVEDTAARL-----QSKGISAAAYHAGLENNVRADVQSKF 281
 Qy 761 MQQLRVVAVTAVFGMLDRPDVRAVLHGLGPPSFESVYQAVGRAGRDQPAHCHLFL- 818
 Db 282 QRDDLQIVVATVAFGMGINKPNVFPVHFDPRIPIETSYQETGRAGRDGLPAEAMLFYDP 341
 Qy 819 -----QFGEDLRELRRH---VHADSTDFLAVKRLV-----QRVFPACTC- 855
 Db 342 ADMAWLRCLKEKQGG-LQDIERHKLNAMGAFAEATCRRLVLLNYFEGRQEGPCGNG 400
 Qy 856 TCTRPPSPQEGAVGGE 871
 Db 401 ICLDPPKQYDGDSTDAQ 416

RESULT 15
 Q8X8N1 ECO57 PRELIMINARY; PRT; 611 AA.
 AC Q8X8N1_07A9D7;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DE 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
 DE ATP-dependent DNA helicase.
 GN Name=recQ; OrderedLocusNames=EC84752, z5343;
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Escherichia*.
 OX NCBI_TaxID=83334;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
 RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:43:16 ; Search time 52 Seconds
(without alignments)
1920.620 Million cell updates/sec

Title: US-09-889-325-4
Perfect score: 6424
Sequence: 1 MERLDVRLQAWERAFRR.....HLSFHALVGLATELLOVAR 1208

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/PCRTUS COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	2	US-09-463-702A-2
2	6424	100.0	1208	2	US-09-699-135-2
3	1883	29.3	361	2	US-09-463-702A-37
4	1883	29.3	361	2	US-09-699-135-37
5	632.5	9.8	607	2	US-08-781-891-75
6	632.5	9.8	607	2	US-09-618-166-75
7	621	9.7	370	1	US-08-559-303B-77
8	621	9.7	370	2	US-09-175-828-77
9	621	9.7	370	2	US-09-753-143-77
10	613.5	9.6	334	2	US-09-463-702A-38
11	613.5	9.6	334	2	US-09-699-135-38
12	606	9.4	619	2	US-09-489-039A-8075
13	591.5	9.2	641	2	US-09-543-681A-4359
14	595	9.1	1394	2	US-09-902-540-16497
15	554.5	8.6	589	2	US-09-134-000C-5710
16	547	8.5	849	2	US-09-252-991A-17953
17	545	8.5	106	2	US-09-463-702A-43
18	545	8.5	106	2	US-09-699-135-43
19	529	8.2	616	2	US-09-107-532A-7258
20	526	8.2	637	2	US-09-134-001C-5658
21	525.5	8.2	1328	2	US-08-781-891-76
22	525.5	8.2	1328	2	US-09-618-166-76
23	521	8.1	473	2	US-09-107-532A-5941
24	506.5	7.9	410	2	US-09-949-016-9327
25	494	7.7	355	2	US-09-902-540-15688
26	487	7.6	393	1	US-08-559-303B-76
27	487	7.6	393	2	US-09-175-828-76

28	487	7.6	393	2	US-09-753-143-76	Sequence 76, Appl
29	485.5	7.6	1401	2	US-09-127-670-6	Sequence 6, Appli
30	481.5	7.5	333	2	US-09-463-702A-39	Sequence 39, Appl
31	481.5	7.5	333	2	US-09-699-135-39	Sequence 39, Appl
32	479.5	7.5	1401	2	US-08-781-891-206	Sequence 206, App
33	479.5	7.5	1401	2	US-09-618-166-206	Sequence 206, App
34	478.5	7.4	330	2	US-09-134-000C-4772	Sequence 4772, Ap
35	476	7.4	1409	2	US-09-949-001-22	Sequence 22, Appl
36	475	7.4	1269	2	US-08-781-891-74	Sequence 74, Appl
37	475	7.4	1269	2	US-09-618-166-74	Sequence 74, Appl
38	475	7.4	1432	2	US-08-781-891-71	Sequence 71, Appl
39	475	7.4	1432	2	US-09-618-166-71	Sequence 71, Appl
40	475	7.4	1432	2	US-09-949-001-16	Sequence 16, Appl
41	473	7.4	879	2	US-09-248-796A-20040	Sequence 20040, A
42	464.5	7.2	386	1	US-08-559-303B-75	Sequence 75, Appl
43	464.5	7.2	386	2	US-09-175-828-75	Sequence 75, Appl
44	464.5	7.2	386	2	US-09-753-143-75	Sequence 75, Appl
45	464.5	7.2	659	2	US-08-781-891-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-09-463-702A-2
; Sequence 2, Application US/09463702A
; Patent No. 6335435
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO, AKIRO
; APPLICANT: KITAO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; FILE OF INVENTION: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/463,702A
; CURRENT FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-463-702A-2

Query Match	100.0%	Score	6424	DB	2	Length	1208
Best Local Similarity	100.0%	Pred. No.	0	Mismatches	0	Indels	0
Matches	1208	Conservative	0				
Qy	1	MERLDVRLQAWERAFRRQRRPSQDDVEAAPEETRALYREYRTLKRTTGOAGGGLR	60				
Db	1	MERLDVRLQAWERAFRRQRRPSQDDVEAAPEETRALYREYRTLKRTTGOAGGGLR	60				
Qy	61	SSESLPAAAEBAEPRCWMGPHLNRAATKSPQTPGSRQSGVDPYQBLKANLKTLOAG	120				
Db	61	SSESLPAAAEBAEPRCWMGPHLNRAATKSPQTPGSRQSGVDPYQBLKANLKTLOAG	120				
Qy	121	PALGHRPWLGRASKASTPKPGTGPVPSFAEKVSDPEPQLPEPQPPRGLHLOASLS	180				
Db	121	PALGHRPWLGRASKASTPKPGTGPVPSFAEKVSDPEPQLPEPQPPRGLHLOASLS	180				
Qy	181	QRLGSLDPGWLQRCHEVPDFLGAPKACRDLGSESOLLIPGESAVLPGAGSGQGPAS	240				
Db	181	QRLGSLDPGWLQRCHEVPDFLGAPKACRDLGSESOLLIPGESAVLPGAGSGQGPAS	240				
Qy	241	AFQEVSIKVGSPQSSSGGKKRRWNNEEPWESPAQVQOESSQAGPPESEGAGVAVEEDPPG	300				
Db	241	AFQEVSIKVGSPQSSSGGKKRRWNNEEPWESPAQVQOESSQAGPPESEGAGVAVEEDPPG	300				

```
QY 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAKAGTAPLHIPPRLARHDRGNYYRLNNKQ 360
D 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAKAGTAPLHIPPRLARHDRGNYYRLNNKQ 360
QY 361 KHYVGRALRSRLRKQAWKQKWKKECFGGGATVTTKESCFLEQFQDHWAACQPRPA 420
D 361 KHYVGRALRSRLRKQAWKQKWKKECFGGGATVTTKESCFLEQFQDHWAACQPRPA 420
QY 421 SEEDTAVGPEPLVPSPQPEVEPSLQPTVPLYSGLSPSGQLAETPAEVFQALQHQ 480
D 421 SEEDTAVGPEPLVPSPQPEVEPSLQPTVPLYSGLSPSGQLAETPAEVFQALQHQ 480
QY 481 FRPQERAVMRILSGISTLLVLPFGAGKSLCYQLPALLYSRSPCLTLVSPILLSMDQ 540
D 481 FRPQERAVMRILSGISTLLVLPFGAGKSLCYQLPALLYSRSPCLTLVSPILLSMDQ 540
QY 541 VSGLPCLKAACIHSQWTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAQLPPVA 600
D 541 VSGLPCLKAACIHSQWTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAQLPPVA 600
QY 601 FACIDEAHCLSQWSHNFRPCYLRCVKLRRMGVHCFLGLTATATRTTASDVQHLAVAE 660
D 601 FACIDEAHCLSQWSHNFRPCYLRCVKLRRMGVHCFLGLTATATRTTASDVQHLAVAE 660
QY 661 EPDLHGPAVPVPTNLHLVSMRDRTDQALLTLQKRFQNLDSIIYCNREDTERIAALL 720
D 661 EPDLHGPAVPVPTNLHLVSMRDRTDQALLTLQKRFQNLDSIIYCNREDTERIAALL 720
QY 721 RTCLHAHVPGSGGRAPKPTTAEAYHAGCSRRERVRQAFMOGOLRVVVAATVAFMGGLDR 780
D 721 RTCLHAHVPGSGGRAPKPTTAEAYHAGCSRRERVRQAFMOGOLRVVVAATVAFMGGLDR 780
QY 781 PDVRAVLHLGLPPSPFESVQAVGRAGRDGPACHLFLQPGEDLRLRHVHADSTDFL 840
D 781 PDVRAVLHLGLPPSPFESVQAVGRAGRDGPACHLFLQPGEDLRLRHVHADSTDFL 840
QY 841 AVKRLVQRVFPACTCTCTRPSPSEQEGAVGGERPVKYPQAEQLSHQAAQPPRRVCMGH 900
D 841 AVKRLVQRVFPACTCTCTRPSPSEQEGAVGGERPVKYPQAEQLSHQAAQPPRRVCMGH 900
QY 901 ERALPIQLTVOALDMPBEATETLLCYLELPHHWELELLATYTHCLNCPGGPAQLQALA 960
D 901 ERALPIQLTVOALDMPBEATETLLCYLELPHHWELELLATYTHCLNCPGGPAQLQALA 960
QY 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPTGV 1020
D 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPTGV 1020
QY 1021 RRGTVLVEFSELAFHLRSPGDLTAEBKQIQCDFLYGRVQARERQALRLRRTFOAFHSV 1080
D 1021 RRGTVLVEFSELAFHLRSPGDLTAEBKQIQCDFLYGRVQARERQALRLRRTFOAFHSV 1080
QY 1081 AFPSGCGLEQDEERSTRKDLGRYFEEBEGQEPGEMDAQEPGQARLQWEDQVR 1140
D 1081 AFPSGCGLEQDEERSTRKDLGRYFEEBEGQEPGEMDAQEPGQARLQWEDQVR 1140
QY 1141 CDIRQFLSLRPEEKFSRAVARIFPHIGSPCYPAQVYQDQRRFWRKYLHLSFHALVGLAT 1200
D 1141 CDIRQFLSLRPEEKFSRAVARIFPHIGSPCYPAQVYQDQRRFWRKYLHLSFHALVGLAT 1200
QY 1201 BELLOVAR 1208
D 1201 BELLOVAR 1208
```

RESULT 2

US-09-699-135-2

; Sequence 2, Application US/09699135

; Patent No. 6472513

; GENERAL INFORMATION:

; APPLICANT: AGENE Research Institute, Co., Ltd.

; APPLICANT: HIRAKI AND ASSOCIATES

; APPLICANT: SHIMAMOTO, AKIRO

```
; APPLICANT: KITAO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE REQ4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/699,135
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-699-135-2
```

Query Match 100.0%; Score 6424; DB 2; Length 1208;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MERLRDVRERLQAWERAFRRQGRRPSSQDDVEAAPETRALYREYRTLKRTTGQAGGLR 60
D 1 MERLRDVRERLQAWERAFRRQGRRPSSQDDVEAAPETRALYREYRTLKRTTGQAGGLR 60
QY 61 SSESLLPAAAEAEPCPCWGHILNRAATKSPQTPGRSRQSGVDPYQORLKANLKGTLQAG 120
D 61 SSESLLPAAAEAEPCPCWGHILNRAATKSPQTPGRSRQSGVDPYQORLKANLKGTLQAG 120
QY 121 PALGRRPWPLGRASSKASTPKPGTGPVPSFAKVSDEPPQLPEPQPRGRLOHLOASLS 180
D 121 PALGRRPWPLGRASSKASTPKPGTGPVPSFAKVSDEPPQLPEPQPRGRLOHLOASLS 180
QY 181 QRLGSLDPCWLQCHSEVPDFLQAPKACRDLGSEESQLLIPGESAVLPGAGSQSGPEAS 240
D 181 QRLGSLDPCWLQCHSEVPDFLQAPKACRDLGSEESQLLIPGESAVLPGAGSQSGPEAS 240
QY 241 AFEVSIIRVSGPQPSGSGGKRRNBPWSPQAQVQOESSQAQPPSPGAGAVAEEDPPG 300
D 241 AFEVSIIRVSGPQPSGSGGKRRNBPWSPQAQVQOESSQAQPPSPGAGAVAEEDPPG 300
QY 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAKAGTAPLHIPPRLARHDRGNYYRLNNKQ 360
D 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAKAGTAPLHIPPRLARHDRGNYYRLNNKQ 360
QY 361 KHYVGRALRSRLRKQAWKQKWKKECFGGGATVTTKESCFLEQFQDHWAACQPRPA 420
D 361 KHYVGRALRSRLRKQAWKQKWKKECFGGGATVTTKESCFLEQFQDHWAACQPRPA 420
QY 421 SEEDTAVGPEPLVPSPQPEVEPSLQPTVPLYSGLSPSGQLAETPAEVFQALQHQ 480
D 421 SEEDTAVGPEPLVPSPQPEVEPSLQPTVPLYSGLSPSGQLAETPAEVFQALQHQ 480
QY 481 FRPQERAVMRILSGISTLLVLPFGAGKSLCYQLPALLYSRSPCLTLVSPILLSMDQ 540
D 481 FRPQERAVMRILSGISTLLVLPFGAGKSLCYQLPALLYSRSPCLTLVSPILLSMDQ 540
QY 541 VSGLPCLKAACIHSQWTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAQLPPVA 600
D 541 VSGLPCLKAACIHSQWTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAQLPPVA 600
QY 601 FACIDEAHCLSQWSHNFRPCYLRCVKLRRMGVHCFLGLTATATRTTASDVQHLAVAE 660
D 601 FACIDEAHCLSQWSHNFRPCYLRCVKLRRMGVHCFLGLTATATRTTASDVQHLAVAE 660
QY 661 EPDLHGPAVPVPTNLHLVSMRDRTDQALLTLQKRFQNLDSIIYCNREDTERIAALL 720
D 661 EPDLHGPAVPVPTNLHLVSMRDRTDQALLTLQKRFQNLDSIIYCNREDTERIAALL 720
QY 721 RTCLHAHVPGSGGRAPKPTTAEAYHAGCSRRERVRQAFMOGOLRVVVAATVAFMGGLDR 780
```

Db 721 RTCLHAAPVPGSGGRAPKTTAEAYHAGCSCRERRVQRAFMQGLRVVVVATVAFGMGLDR 780
 Qy 781 PDVRAVLHLGLPPSPFESVQVAGRAGRDGQPAHCHLFLQPGEDLRELRRHVHADSTDFL 840
 Db 781 PDVRAVLHLGLPPSPFESVQVAGRAGRDGQPAHCHLFLQPGEDLRELRRHVHADSTDFL 840
 Qy 841 AVKGLVQRPACTCTCTRPSPBEGAVGGRRPVKYPQPAEQLSHQAAPGPRVCMGH 900
 Db 841 AVKGLVQRPACTCTCTRPSPBEGAVGGRRPVKYPQPAEQLSHQAAPGPRVCMGH 900
 Qy 901 BRALPIQLTVQALDMPERAIETLLCYLHLPWHLELLATTYTHCRNLNCPGPGPAQLQALA 960
 Db 901 BRALPIQLTVQALDMPERAIETLLCYLHLPWHLELLATTYTHCRNLNCPGPGPAQLQALA 960
 Qy 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDEHPTGV 1020
 Db 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDEHPTGV 1020
 Qy 1021 RRGTVLVEFSELAFHLRSPGDLTAEEKDQICDFLYGRVQARERQALRLRTTQAFHSV 1080
 Db 1021 RRGTVLVEFSELAFHLRSPGDLTAEEKDQICDFLYGRVQARERQALRLRTTQAFHSV 1080
 Qy 1081 APPSCGCLQOODEERSTLXDLGRYFEEBEGQEPGMEQAQGPQOARLQWEDQVR 1140
 Db 1081 APPSCGCLQOODEERSTLXDLGRYFEEBEGQEPGMEQAQGPQOARLQWEDQVR 1140
 Qy 1141 CDIROFLSLRPEEKFSRAVARIFHGIGSPCYPAQVYGQDRFRFWRKYLHLSFHALVGLAT 1200
 Db 1141 CDIROFLSLRPEEKFSRAVARIFHGIGSPCYPAQVYGQDRFRFWRKYLHLSFHALVGLAT 1200
 Qy 1201 BELLQVAR 1208
 Db 1201 BELLQVAR 1208

RESULT 3

US-09-463-702A-37
 ; Sequence 37, Application US/09463702A
 ; Patent No. 6335435
 ; GENERAL INFORMATION:
 ; APPLICANT: AGENE Research Institute, Co., Ltd.
 ; APPLICANT: HIRAKI AND ASSOCIATES
 ; APPLICANT: SHIMAMOTO, AKIRO
 ; APPLICANT: KITAO, SAORI
 ; APPLICANT: FURUICHI, YASUHIRO
 ; TITLE OF INVENTION: HUMAN GENE REQ4 ENCODING HELICASE
 ; FILE REFERENCE: HIRAI150
 ; CURRENT APPLICATION NUMBER: US/09/463, 702A
 ; PRIOR FILING DATE: 2000-01-24
 ; PRIOR APPLICATION NUMBER: PCT/JP98/03114
 ; PRIOR FILING DATE: 1998-07-10
 ; PRIOR APPLICATION NUMBER: JAPAN 9/200387
 ; PRIOR FILING DATE: 1997-07-25
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 37
 ; LENGTH: 361
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-463-702A-37

Query Match 29.3%; Score 1883; DB 2; Length 361;
 Best Local Similarity 100.0%; Pred. No. 2.4e-147;
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 471 QALEQLGHOAFRPGQERAVNRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVV 530
 Db 1 QALEQLGHOAFRPGQERAVNRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVV 60
 Qy 531 SPLSLMDQVSGLPCLKAACHSGMTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGL 590
 Db 61 SPLSLMDQVSGLPCLKAACHSGMTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGL 120

Qy 591 PPAQLPPVAFACIDEAHCLSQSHNFRPCYLVRCKVLRRMGVHCFGLGTATATRTAS 650
 Db 121 PPAQLPPVAFACIDEAHCLSQSHNFRPCYLVRCKVLRRMGVHCFGLGTATATRTAS 180
 Qy 651 DVAQHHLAVAEEDPOLHGPAPVPTNLHLSVSMRDRTDQALLTLQGRFQNLDSIIYCNRR 710
 Db 181 DVAQHHLAVAEEDPOLHGPAPVPTNLHLSVSMRDRTDQALLTLQGRFQNLDSIIYCNRR 240
 Qy 711 EDTERIAALLRCTLHAAWPGSGGRAPKTTAEAYHAGCSCRERRVQRAFMQGLRVVVA 770
 Db 241 EDTERIAALLRCTLHAAWPGSGGRAPKTTAEAYHAGCSCRERRVQRAFMQGLRVVVA 300
 Qy 771 TVAFGMGLDRPDVRAVLHLGLPPSFESVQVAGRAGRDGQPAHCHLFLQPGEDLRELRR 830
 Db 301 TVAFGMGLDRPDVRAVLHLGLPPSFESVQVAGRAGRDGQPAHCHLFLQPGEDLRELRR 360
 Qy 831 H 831
 Db 361 H 361

RESULT 4

US-09-699-135-37
 ; Sequence 37, Application US/09699135
 ; Patent No. 6472513
 ; GENERAL INFORMATION:
 ; APPLICANT: AGENE Research Institute, Co., Ltd.
 ; APPLICANT: HIRAKI AND ASSOCIATES
 ; APPLICANT: SHIMAMOTO, AKIRO
 ; APPLICANT: KITAO, SAORI
 ; APPLICANT: FURUICHI, YASUHIRO
 ; TITLE OF INVENTION: HUMAN GENE REQ4 ENCODING HELICASE
 ; FILE REFERENCE: HIRAI150
 ; CURRENT APPLICATION NUMBER: US/09/699,135
 ; PRIOR FILING DATE: 2000-10-27
 ; PRIOR APPLICATION NUMBER: US/09/463, 702A
 ; PRIOR FILING DATE: 2000-01-24
 ; PRIOR APPLICATION NUMBER: PCT/JP98/03114
 ; PRIOR FILING DATE: 1998-07-10
 ; PRIOR APPLICATION NUMBER: JAPAN 9/200387
 ; PRIOR FILING DATE: 1997-07-25
 ; NUMBER OF SEQ ID NOS: 44
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 37
 ; LENGTH: 361
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-699-135-37

Query Match 29.3%; Score 1883; DB 2; Length 361;
 Best Local Similarity 100.0%; Pred. No. 2.4e-147;
 Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 471 QALEQLGHOAFRPGQERAVNRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVV 530
 Db 1 QALEQLGHOAFRPGQERAVNRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVV 60
 Qy 531 SPLSLMDQVSGLPCLKAACHSGMTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGL 590
 Db 61 SPLSLMDQVSGLPCLKAACHSGMTRKQRESVLQKIRAAQVHVLMLTPEALVGAGGL 120
 Qy 591 PPAQLPPVAFACIDEAHCLSQSHNFRPCYLVRCKVLRRMGVHCFGLGTATATRTAS 650
 Db 121 PPAQLPPVAFACIDEAHCLSQSHNFRPCYLVRCKVLRRMGVHCFGLGTATATRTAS 180
 Qy 651 DVAQHHLAVAEEDPOLHGPAPVPTNLHLSVSMRDRTDQALLTLQGRFQNLDSIIYCNRR 710
 Db 181 DVAQHHLAVAEEDPOLHGPAPVPTNLHLSVSMRDRTDQALLTLQGRFQNLDSIIYCNRR 240
 Qy 711 EDTERIAALLRCTLHAAWPGSGGRAPKTTAEAYHAGCSCRERRVQRAFMQGLRVVVA 770
 Db 241 EDTERIAALLRCTLHAAWPGSGGRAPKTTAEAYHAGCSCRERRVQRAFMQGLRVVVA 300

[illegible]

```

704  ---IYYCNREDTERIAALIRLTC LHAAYVPGSGGRAPKTTAEAYHAGWCSERRERVRQAF 760
236  KSGIYYCNSRAKVEDTAAAL-----QSGKISAAAYHAGLENNVRADVQEKF 281
761  MQOGLRVVVATVAFGMLGDPDYAVLHGLPPSFESYQAVGRAGRDGQPAHCHLFL-- 818
282  QRDQLQIVVATVAFGMLGDKPNVRFVVFHFDIPRIEYSYQETGRAGRDGLPAEAMLFYDP 341
819  -----OPQGBDLRELRRH---VHADSTDFLAVKRLV-----QRVFPACTC- 855
342  ADMAWLRLCLEEPQGO-LQDIERHKLNAMGAFABEQTCRLRLVLLNYFEGEGQEPGNCND 400
856  TCTRPPSEBQAVGGE 871
401  ICLDPPKQYDGDSTAQ 416

RESULT 6
US-09-618-166-75
; Sequence 75, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; Yu, Chang-En
; Oshima, Junko
; Mulligan, John T.
; Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/618,166
; FILING DATE: 17-Jul-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.419C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-618-166-75

```

```

Db 74 ISLMKQVQDQVQANGVAAACLNSTQREQQLEVMVTCGTGQIRLLYIAPERLMLDNFLEH 133
Qy 593 AAQLPPVAFACIDEAHCLSQSHNFRPCYLKRVCKVLRMRMGVHCFGLGTATATARTASDV 652
Db 134 LAHNVPVLLA-VDEAHCSIQMGHDFRPEYALGQ-LRQRPPTLPFMALTATADDTTRQDI 191
Qy 653 AQHLAVAEEDPLHGPAPVPTNLHLVSVMDRDTDQALLTLLOGKRFQNLDSI----- 703
Db 192 VRLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFKPLDQLMRYVQORQ 235
Qy 704 ---IICNRREDTERIAALLRTCLHAAWVPGSGGRAPKTTAEAYHAGMCSRERRVORAF 760
Db 236 KSGIYCNSRAKVEDTAAAL-----QSKGISAAAYHAGLENNVRADVQKPF 281
Qy 761 MQQLRVVAVATVAFMGDLRDPDVRVAVLHGLPPSFESYVQAVGRAGRDGQPAHCHLFL-- 818
Db 282 QRDDLQIVVATVAFMGINKENVRVVFHFDIPRNIESYVQETGRAGRDGLPAEAMLFYDP 341
Qy 819 -----QPGEDLRLRH---VHADSTDFLAVKELV-----QRFVPACTC- 855
Db 342 ADMAWLRCLCEKPGQ-LQDIERHKLNAMGAFEAOTCRRLVLLNYPGEGRQBPFCNGCD 400
Qy 856 TCTRPSPQEGAVGGE 871
Db 401 ICLDPPKQYDGSDDAQ 416

```

RESULT 7

```

US-08-559-303B-77
; Sequence 77, Application US/08559303B
; Patent No. 5824501
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,303B
; FILING DATE: NOVEMBER 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:

```

```

; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; US-08-559-303B-77
;
; Query Match 9.7%; Score 621; DB 1; Length 370;
; Best Local Similarity 40.0%; Pred. No. 7.1e-43;
; Matches 154; Conservative 57; Mismatches 110; Indels 64; Gaps 12;
;
; Qy 474 EQLGHQAFRRGQRRVNRILSGISTLLVLTGAGKSLCYQLPALLYRRSPCLTLVVSP 533
; Db 4 ETEGYOQFRPQEBIIDVLISGRDCLVVMPTGGKSLCYQIPALLNG-----LTVVVSPL 59
;
; Qy 534 LSLMDDDOVSGLPP-CLKAACIHSQMTKQRESVLQKIRAAQVHVLMLTPEALVCGAGLPP 592
; Db 60 ISLMKQVQDQVQANGVAAACLNSTQREQQLEVMVTCGTGQIRLLYIAPERLMLDNFLEH 119
; Qy 593 AAQLPPVAFACIDEAHCLSQSHNFRPCYLKRVCKVLRMRMGVHCFGLGTATATARTASDV 652
; Db 120 LAHNVPVLLA-VDEAHCSIQMGHDFRPEYALGQ-LRQRPPTLPFMALTATADDTTRQDI 177
;
; Qy 653 AQHLAVAEEDPLHGPAPVPTNLHLVSVMDRDTDQALLTLLOGKRFQNLDSI----- 703
; Db 178 VRLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFKPLDQLMRYVQORQ 221
; Qy 704 ---IICNRREDTERIAALLRTCLHAAWVPGSGGRAPKTTAEAYHAGMCSRERRVORAF 760
; Db 222 KSGIYCNSRAKVEDTAAAL-----QSKGISAAAYHAGLENNVRADVQKPF 267
; Qy 761 MQQLRVVAVATVAFMGDLRDPDVRVAVLHGLPPSFESYVQAVGRAGRDGQPAHCHLFL-- 818
; Db 268 QRDDLQIVVATVAFMGINKENVRVVFHFDIPRNIESYVQETGRAGRDGLPAEAMLFYDP 327
;
; Qy 819 -----QPGEDLRLRH 831
; Db 328 ADMAWLRCLCEKPGQ-LQDIERH 351
;
; RESULT 8
; US-09-175-828-77
; Sequence 77, Application US/09175828
; Patent No. 6221643
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; APPLICANT: GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/175,828
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,303
; FILING DATE: NOVEMBER 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766

```

```
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
;
US-09-175-828-77

Query Match          9.7%; Score 621; DB 2; Length 370;
Best Local Similarity 40.0%; Pred. No. 7.1e-43;
Matches 154; Conservative 57; Mismatches 110; Indels 64; Gaps 12;

QY 474 EQLGHOAFRPGQRAVNRILSGISTLLVPTGAGKSLCYQLPALLYRRSPCLTLVVSPL 533
DB 4 ETFGYQFRPGQEEIITVLSGRDCLVVMPTGGKSLCYQIPALLNG----LTVVVSPL 59

QY 534 LSLMDDQVSGLPP-CLKAACIHSGMTRKQRESVLQKIRAAQVHVLMTPALVAGAGLPP 592
DB 60 ISLMKDQVDQLQANGVAACLNSTQREQLVMTGCTGTGQIRLLYIAPERLMLDNFLEH 119

QY 593 AAQLPPVAFACIDAHCLSSQSHNFRPCYLKRVCKLRRMGVHCFGLGTATATRTASDV 652
DB 120 LAHWNPVLLA-VDEAHCSQWGHDFRPEYALGQ-LRQRPPTLPFMALTATADTTTQDI 177

QY 653 AQHLAVAEEDLHGPAVPPTNLHLSVSMRDTQALLTLQKGFQNLDSI----- 703
DB 178 VRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFKPLDQLMRVYVQSRG 221

QY 704 ---IICNRRDTERIAALLRTCLHAAWVPGSGRPAKTTAEAYHAGMCSRERRVORAF 760
DB 222 KSGIIYCNRAKVEDTAAAL-----QSKGISAAAYHAGLENNVRADVQEKF 267

QY 761 MQQLRVVATVAFGMGLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGQPAHCHFL-- 818
DB 268 QRDDLQIVVATVAFGMGINKPNRVFVHFDPINIESYVQETGRAGRDGLPAEAMLFYDP 327

QY 819 -----QOQEDLRELRRH 831
DB 328 ADMAWLRRCLEEKPOGO-LQDIERH 351

RESULT 9
US-09-753-143-77
; Sequence 77, Application US/09753143
; Patent No. 6838240
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF BLOOM'S SYNDROME
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
; STREET: 90 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE DISKETTE
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/753,143
; FILING DATE: 02-Jan-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/175,828
; FILING DATE: 1998-10-20
; ATTORNEY/AGENT INFORMATION:
; NAME: ELIZABETH A. BOGOSIAN
; REGISTRATION NUMBER: 39,911
; REFERENCE/DOCKET NUMBER: 63475/65
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-5995
; TELEFAX: (212) 286-0854 or 286-0082
; TELEX: TWX 710-581-4766
; INFORMATION FOR SEQ ID NO: 77:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 370
; TYPE: AMINO ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: <Unknown>
; DESCRIPTION: OTHER NUCLEIC ACID
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 77:
;
US-09-753-143-77

Query Match          9.7%; Score 621; DB 2; Length 370;
Best Local Similarity 40.0%; Pred. No. 7.1e-43;
Matches 154; Conservative 57; Mismatches 110; Indels 64; Gaps 12;

QY 474 EQLGHOAFRPGQRAVNRILSGISTLLVPTGAGKSLCYQLPALLYRRSPCLTLVVSPL 533
DB 4 ETFGYQFRPGQEEIITVLSGRDCLVVMPTGGKSLCYQIPALLNG----LTVVVSPL 59

QY 534 LSLMDDQVSGLPP-CLKAACIHSGMTRKQRESVLQKIRAAQVHVLMTPALVAGAGLPP 592
DB 60 ISLMKDQVDQLQANGVAACLNSTQREQLVMTGCTGTGQIRLLYIAPERLMLDNFLEH 119

QY 593 AAQLPPVAFACIDAHCLSSQSHNFRPCYLKRVCKLRRMGVHCFGLGTATATRTASDV 652
DB 120 LAHWNPVLLA-VDEAHCSQWGHDFRPEYALGQ-LRQRPPTLPFMALTATADTTTQDI 177

QY 653 AQHLAVAEEDLHGPAVPPTNLHLSVSMRDTQALLTLQKGFQNLDSI----- 703
DB 178 VRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFKPLDQLMRVYVQSRG 221

QY 704 ---IICNRRDTERIAALLRTCLHAAWVPGSGRPAKTTAEAYHAGMCSRERRVORAF 760
DB 222 KSGIIYCNRAKVEDTAAAL-----QSKGISAAAYHAGLENNVRADVQEKF 267

QY 761 MQQLRVVATVAFGMGLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGQPAHCHFL-- 818
DB 268 QRDDLQIVVATVAFGMGINKPNRVFVHFDPINIESYVQETGRAGRDGLPAEAMLFYDP 327

QY 819 -----QOQEDLRELRRH 831
DB 328 ADMAWLRRCLEEKPOGO-LQDIERH 351

RESULT 10
US-09-463-702A-38
; Sequence 38, Application US/09463702A
; Patent No. 6335435
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO, AKIRO
; APPLICANT: KITAO, SAORI
```

; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/463,702A
; CURRENT FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-463-702A-38

Query Match 9.6%; Score 613.5; DB 2; Length 334;
Best Local Similarity 41.4%; Pred. No. 2.6e-42;
Matches 152; Conservative 53; Mismatches 111; Indels 51; Gaps 11;
QY 477 GHAQFRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVVSPLLSL 536
Db 2 GYQFRPQGEIITDVLGRDCLVVMPTGGKSLCYQIPALLNG---LTVVVSPLLSL 57
QY 537 MDDQVSGLPP-CLKAACIHSMTKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAAQ 595
Db 58 MKDQVDQLQANGVAAACLNSTQREQLVMTGCTGQIRLLYIAPERLMLDNFLEHLAH 117
QY 596 LPPVAFACIDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGTATATRTTASDVAAH 655
Db 118 WNPVLLA-VDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGTATATRTTASDVAAH 175
QY 656 LAVAEPDLHGPAVPPTNLHLSVSMRDYDQALLTLQKRFQNLDSI----- 703
Db 176 LG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPLDQLMRYVQEQGKSG 219
QY 704 IYCNREDTERIAALLRTCLHAADVPGSGRPAKTTAAAYHAGMCSRERRRVQAFMQG 763
Db 220 IYCNRAKVDTAARL-----OSKGISAAAYHAGLENNVRADVQEKFOR 265
QY 764 QLRVVVATVAFGMGLDRPDVRAVLHGLPPSFESVQAVGRAGDQPAHCHLFLQPOGE 823
Db 266 DLQIVVATVAFGMGINKENVRVVFHDPINIESYQETGRAGRDGLPAEAMLFYDP--A 323
QY 824 DLRELRR 830
Db 324 DMAILRR 330

RESULT 11
US-09-699-135-38
; Sequence 38, Application US/09699135
; Patent No. 6472513
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO, AKIRO
; APPLICANT: KITAO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/699,135
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38

; LENGTH: 334
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-699-135-38
Query Match 9.6%; Score 613.5; DB 2; Length 334;
Best Local Similarity 41.4%; Pred. No. 2.6e-42;
Matches 152; Conservative 53; Mismatches 111; Indels 51; Gaps 11;
QY 477 GHAQFRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVVSPLLSL 536
Db 2 GYQFRPQGEIITDVLGRDCLVVMPTGGKSLCYQIPALLNG---LTVVVSPLLSL 57
QY 537 MDDQVSGLPP-CLKAACIHSMTKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAAQ 595
Db 58 MKDQVDQLQANGVAAACLNSTQREQLVMTGCTGQIRLLYIAPERLMLDNFLEHLAH 117
QY 596 LPPVAFACIDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGTATATRTTASDVAAH 655
Db 118 WNPVLLA-VDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGTATATRTTASDVAAH 175
QY 656 LAVAEPDLHGPAVPPTNLHLSVSMRDYDQALLTLQKRFQNLDSI----- 703
Db 176 LG-----LNDPL-----IQIS-SFDRPNIRYMLM-----EKFPLDQLMRYVQEQGKSG 219
QY 704 IYCNREDTERIAALLRTCLHAADVPGSGRPAKTTAAAYHAGMCSRERRRVQAFMQG 763
Db 220 IYCNRAKVDTAARL-----OSKGISAAAYHAGLENNVRADVQEKFOR 265
QY 764 QLRVVVATVAFGMGLDRPDVRAVLHGLPPSFESVQAVGRAGDQPAHCHLFLQPOGE 823
Db 266 DLQIVVATVAFGMGINKENVRVVFHDPINIESYQETGRAGRDGLPAEAMLFYDP--A 323
QY 824 DLRELRR 830
Db 324 DMAILRR 330
RESULT 12
US-09-489-039A-8075
; Sequence 8075, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8075
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8075
Query Match 9.4%; Score 606; DB 2; Length 619;
Best Local Similarity 35.5%; Pred. No. 2.7e-41;
Matches 169; Conservative 70; Mismatches 157; Indels 80; Gaps 18;
QY 474 EQLGHAQFRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVVSPL 533
Db 30 ETFGYYQFRPQGEIITDVLGRDCLVVMPTGGKSLCYQIPALLNG---LTVVVSPL 85
QY 534 LSLMDDQVSG-LPPCLKAACIHSMTKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPP 592
Db 86 ISLMKDDQVDQLANGVAAACLNSTQREQLVMTGCTGQIRLLYIAPERLMLDNFLEHL 145
QY 593 AAQLPPVAFACIDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGTATATRTTASDV 652
Db 146 LANW-NLAWLAVDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGTATATRTTASDV 203

Db 637 VNRARTPASAAGTG-----QRPVRRAPTAPALRPLK----- 671
Qy 473 LEQLGHQAFRQGERAVNRILSGISLTLVPTGAGKSLCYQLPALLYRRSPCLTLVVSP 532
Db 672 -EAFGSDFRYPQEAACVACRAATAGEDLLVMPGTGAGKSLCYQLPGLARAG----TTLVVSP 726
Qy 533 LLSLMDDDVSGLPPC-LXAACIHSGMTKQRESVLQKTRAAQVHVLMLTPEALVAGAGLP 591
Db 727 LIALMEDQVRLQSLGFAADRIHSGDRAMSRQVCADYLEDRDLDFLTAPERLGVPGFVE 786
Qy 592 PAALQPPVAFACIDEAHCLQSHNSHFRPCYLKRVCKVLRRMGV---HCFLGLTATATRT 648
Db 787 FLARTP-ALTAVDAHACISQWHDPRDY----RLGARPLRLRPAPVVALTATATPDV 841
Qy 649 ASDVAQHLAVAEEDLHGAPVP-----TNLHLSVSMRDRTD-----QALLTLQ 694
Db 842 QRDIVQQLG-----LQPGGKARTFIHGRRTNTAIEV---RELNPGARGDAIQLLED 892
Qy 695 KRFQNLDSIIYCNRETERIAALLRTCLHAAWVPGSGGRAPKTTAEAYHAGMCSRERR 754
Db 893 EENR---PAIVYAATRKAEQLADQL-----AGEFP---AAAYHAGLQPSERD 934
Qy 755 RVORAFMOGLRVVATVAFGMGLDRPDVRAVLHLGLPPSPESYVQAVGRAGRDGQPAHC 814
Db 935 RVQAEFLRGSLEVIATVAFGMGIDKADVTVIHAALPASLEGYYQELGRAGRDGKPSRA 994
Qy 815 ---HLFLOQGEDLRELRRHVHADSTDFLAVKRLVQRFVFPACTCTCTPPSEQEGAVGE 871
Db 995 VLLHSYIDRRTHFPHRRDYPEA-----YVLERLFKS-TAPQLRPFKAVLQGRVRGD 1044
Qy 872 RPVKYPPQEAQOL 885
Db 1045 ---PEVFDKALEQL 1055

RESULT 15
US-09-134-000C-5710
; Sequence 5710, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5710
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5710

Query Match 8.6%; Score 554.5; DB 2; Length 589;
Best Local Similarity 34.0%; Pred. No. 4.6e-37;
Matches 143; Conservative 67; Mismatches 147; Indels 63; Gaps 10;
Qy 472 ALEQL-----GHAQFRPQGERAVMRILSGISLTLVPTGAGKSLCYQLPALLYRRSPCL 526
Db 3 ALQELLKDTFGYDDFRPGQETIIRHVRQENVLGMPTGGKSIQYQLPALLDN----L 58
Qy 527 TLVSPSLSLMDQVS-----GLPCLKAACHSGMTRKQRESVLQKTRAAQVHVLMLTLP 581
Db 59 TLVISPLISLMKDQVDALNMGIP-----ATYINSTISYQEMNHRITQLAVNKEVKLLYVAP 114
Qy 582 EALVGAGGLPPAAQLPVAFACIDEAHCLSQWNSHFRPCYLKRVCKVLRRMGVHCFGLT 641
Db 115 ERLSYDFQOQLTHV-PIDLLAVDEAHACISQWHDPRPSYLRILAEIIDFQQOPTVIALT 173

Qy 642 ATATRTASDVAQHLAVAEEDLHGAPVPPTNLHLSVSMRDRTDQALLTLTLOGKRQFONLD 701
Db 174 ATATPQVAEDIVKQLRIPSENEIK-TGFARENLSFQVVKDQNRDVFLELYLKMTGQ--- 229
Qy 702 SIIYCNRRREDTERIAALLRTCLHAAWVPGSGGRAPKTTAEAYHAGMCSRERRRVQRAFM 761
Db 230 SGIIYASTRKEVERIVHLL-----ESKKIAAGMYHGGMSEQLRSENQEAFL 275
Qy 762 QGQLRVVATVATGMLDRPDVRAVLHLGLPPSPESYVQAVGRAGRDGQPAHCHLFLQPO 821
Db 276 YDQVQVMVATNAFGMGINKSNVRPVIHAQVFGNIESYYQAGRAGRDGLPSDAVIMFAPQ 335
Qy 822 GED-----LRELRRHVHADSTDFLAVKRLVQRFVFPACTCTCTR 859
Db 336 DLQIQYFIEQSEMTIDYKQKEYLKLREMSQYANAQ-----MCLQKYLIRYFGEETDCGR 391

Search completed: December 27, 2005, 21:55:16
Job time : 55 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:53:22 ; Search time 178.4 Seconds
(without alignments)
2829.248 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLDVRELRQAWERAFRR.....HLSPHALVGLATELLQVAR 1208

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	3	US-09-889-325-4
2	6424	100.0	1208	5	US-10-723-860-4382
3	6424	100.0	1208	5	US-10-882-403-4
4	6424	100.0	1208	5	US-10-756-149-5768
5	3944	61.4	1216	6	US-11-084-955-2
6	1723.5	26.8	1579	6	US-11-097-143-16743
7	765	11.9	874	4	US-10-437-963-121617
8	632.5	9.8	607	4	US-10-374-077-75
9	631.5	9.8	610	3	US-09-815-242-10414
10	631.5	9.8	610	4	US-10-282-122A-43357
11	621	9.7	370	3	US-09-753-143-77
12	619	9.6	609	4	US-10-282-122A-56210
13	609.5	9.5	609	4	US-10-282-122A-75783
14	609.5	9.5	615	3	US-09-815-242-13747
15	607.5	9.5	615	4	US-10-282-122A-74973
16	606	9.4	608	4	US-10-282-122A-59929
17	593	9.2	610	4	US-10-282-122A-78025
18	589.5	9.2	600	4	US-10-282-122A-69214
19	581.5	9.1	620	4	US-10-282-122A-76978
20	578.5	9.0	632	4	US-10-282-122A-67293
21	578	9.0	619	3	US-09-815-242-11085
22	578	9.0	619	4	US-10-282-122A-58271
23	554.5	8.6	589	4	US-10-282-122A-42447
24	552	8.6	590	4	US-10-282-122A-60691
25	549	8.5	597	4	US-10-282-122A-72864
26	547.5	8.5	589	3	US-09-815-242-10803
27	545	8.5	580	3	US-09-815-242-4959

28	545	8.5	712	3	US-09-815-242-11915	Sequence 11915, A
29	545	8.5	712	4	US-10-282-122A-66476	Sequence 66476, A
30	544.5	8.5	390	4	US-10-282-122A-61441	Sequence 61441, A
31	544	8.5	341	4	US-10-369-493-8990	Sequence 8990, Ap
32	542.5	8.4	1955	4	US-10-293-504-3	Sequence 3, Appli
33	541	8.4	588	4	US-10-437-963-141345	Sequence 141345, A
34	528	8.2	590	4	US-10-282-122A-67618	Sequence 57618, A
35	526	8.2	592	4	US-10-282-122A-70676	Sequence 70676, A
36	526	8.2	637	4	US-10-724-972A-7447	Sequence 7447, Ap
37	525.5	8.2	1328	4	US-10-374-077-76	Sequence 76, Appl
38	524.5	8.2	658	4	US-10-156-761-12385	Sequence 12385, A
39	519.5	8.1	618	4	US-10-425-115-256333	Sequence 256333, A
40	518.5	8.1	715	4	US-10-282-122A-67521	Sequence 67521, A
41	517	8.0	653	4	US-10-425-114-59459	Sequence 59459, A
42	515	8.0	827	4	US-10-282-122A-53151	Sequence 53151, A
43	514	8.0	593	4	US-10-282-122A-70264	Sequence 70264, A
44	513	8.0	593	3	US-09-815-242-12306	Sequence 12306, A
45	511	8.0	601	4	US-10-282-122A-48328	Sequence 48328, A

ALIGNMENTS

RESULT 1
US-09-889-325-4
; Sequence 4, Application US/09889325
; Publication No. US20040224312A1
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; TITLE OF INVENTION: Gene causative of Rothmund-Thomson syndrome
; TITLE OF INVENTION: and its gene product
; FILE REFERENCE: A1-003PCT
; CURRENT APPLICATION NUMBER: US/09/889,325
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: JP 1999-11218
; PRIOR FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-889-325-4

Query Match	100.0%	Score	6424	DB	3	Length	1208
Best Local Similarity	100.0%	Pred. No.	0				
Matches	1208	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MERLDVRELRQAWERAFRRQRPSODDVEAAPEETRALYREYRTLRKTGGAGGLR	60				
Db	1	MERLDVRELRQAWERAFRRQRPSODDVEAAPEETRALYREYRTLRKTGGAGGLR	60				
Qy	61	SSESLPAAAEAEPRCWMGPHLNRAATKSPOPTGSRSGSVDPYQGRLLKANLKTLOAG	120				
Db	61	SSESLPAAAEAEPRCWMGPHLNRAATKSPOPTGSRSGSVDPYQGRLLKANLKTLOAG	120				
Qy	121	PALGRRPPLGRASSKASTPKPGTGPVPSFAEKVSDPPQLPEPQPPRGLQHLQASLS	180				
Db	121	PALGRRPPLGRASSKASTPKPGTGPVPSFAEKVSDPPQLPEPQPPRGLQHLQASLS	180				
Qy	181	QRLGSLDPGWLORCHSEYVDFLGAPKACRDLGSEESOLLIPGESAVLPGAGSGOPEAS	240				
Db	181	QRLGSLDPGWLORCHSEYVDFLGAPKACRDLGSEESOLLIPGESAVLPGAGSGOPEAS	240				
Qy	241	AFQEVSIKVGSPQSSSGCKRRWNEEPWESPAQVQOQSSQAGPPSEGAGAVAEEDPPG	300				
Db	241	AFQEVSIKVGSPQSSSGCKRRWNEEPWESPAQVQOQSSQAGPPSEGAGAVAEEDPPG	300				
Qy	301	EPVQAPQPCSSPSNPRYHGLSPSSQARAGKAEGTAPLHIFPRLARHDRGNVRLNMKQ	360				
Db	301	EPVQAPQPCSSPSNPRYHGLSPSSQARAGKAEGTAPLHIFPRLARHDRGNVRLNMKQ	360				
Qy	361	KHYVGRALRSLLRKQAWKQWRKKGCFGGGATVTTKSCFLNEQPDHWAACQPRPA	420				

Db 361 KHYVGRALRRLRLKQAWKRWKKGCFGGGATVTTKSCFLNEQFQDHWAAQCRPPA 420
Qy 421 SEEDTDAVGPEPLVPSPQVPEVPSLDPTVLPLSLGSPGQLAETPAEVFOALQOLGHOA 480
Db 421 SEEDTDAVGPEPLVPSPQVPEVPSLDPTVLPLSLGSPGQLAETPAEVFOALQOLGHOA 480
Qy 481 FRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVWSPLLSLMDDQ 540
Db 481 FRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVWSPLLSLMDDQ 540
Qy 541 VSGLPCLKAACIHSWTRKQRESVLQKIRAAOVHVLMTPEALVAGGLPPAAQLPPVA 600
Db 541 VSGLPCLKAACIHSWTRKQRESVLQKIRAAOVHVLMTPEALVAGGLPPAAQLPPVA 600
Qy 601 FACIDEAHCLSQWSHNFRPCYLKVRCKLREMGVHCFGLGTATATRTASDVAAHVAE 660
Db 601 FACIDEAHCLSQWSHNFRPCYLKVRCKLREMGVHCFGLGTATATRTASDVAAHVAE 660
Qy 661 EPDLHGPAVPVPTNLHLVSMDRDTDQALLTLLOGKRFQNLDSIIYCNRRREDTERIAALL 720
Db 661 EPDLHGPAVPVPTNLHLVSMDRDTDQALLTLLOGKRFQNLDSIIYCNRRREDTERIAALL 720
Qy 721 RTCLHAAWPGSGGRAPKTTAEAYHAGMCSRERRVQRAFMQGLRVVVAATVAFGMGLDR 780
Db 721 RTCLHAAWPGSGGRAPKTTAEAYHAGMCSRERRVQRAFMQGLRVVVAATVAFGMGLDR 780
Qy 781 PDVRAVLHLGLPPSPFESYVQAVGRAGDQGAHCHLFLQPGEDLRELRRHVHADSTDFL 840
Db 781 PDVRAVLHLGLPPSPFESYVQAVGRAGDQGAHCHLFLQPGEDLRELRRHVHADSTDFL 840
Qy 841 AVKRLVQRVFPACTCTCTRPPSEQEGAVGGERPVKYPQAEQLSHQAAPGPRRVCMGH 900
Db 841 AVKRLVQRVFPACTCTCTRPPSEQEGAVGGERPVKYPQAEQLSHQAAPGPRRVCMGH 900
Qy 901 ERALPIQLTVQALDMPAEIATILCYLELHPHWELELLATYTHCRNLNCGPGPAQLQALA 960
Db 901 ERALPIQLTVQALDMPAEIATILCYLELHPHWELELLATYTHCRNLNCGPGPAQLQALA 960
Qy 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTV 1020
Db 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTV 1020
Qy 1021 RRGTVLVFSELAFLHRLSPGDLTAEKQDQICDFLYGRVQARERQALRLRTFQA FHSV 1080
Db 1021 RRGTVLVFSELAFLHRLSPGDLTAEKQDQICDFLYGRVQARERQALRLRTFQA FHSV 1080
Qy 1081 APPSCGCGLEQDDEERSTRLLKDLGRYFEEBEGQEPGGMEDAQGPFGQARLQWEDQVR 1140
Db 1081 APPSCGCGLEQDDEERSTRLLKDLGRYFEEBEGQEPGGMEDAQGPFGQARLQWEDQVR 1140
Qy 1141 CDIRQFLSLRPEEKFSFRAVARIIPHGIGSPCYPAQVYQDQRRFWRKYLHLHSFHALVGLAT 1200
Db 1141 CDIRQFLSLRPEEKFSFRAVARIIPHGIGSPCYPAQVYQDQRRFWRKYLHLHSFHALVGLAT 1200
Qy 1201 BELLQVAR 1208
Db 1201 BELLQVAR 1208

RESULT 2
US-10-723-860-4382
; Sequence 4382, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26

; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4382
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-4382

Query Match 100.0%; Score 6424; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MERLURVRELOQAWERAFRRQRRRSQDDVEAAPETRALYREYRTLKRTTGQAGGLR 60
Db 1 MERLURVRELOQAWERAFRRQRRRSQDDVEAAPETRALYREYRTLKRTTGQAGGLR 60
Qy 61 SSESLLPAAAEAEPEPCWGPFLNRAATKSPQTPGRSRQSGVDPYQRLKANLKGTLQAG 120
Db 61 SSESLLPAAAEAEPEPCWGPFLNRAATKSPQTPGRSRQSGVDPYQRLKANLKGTLQAG 120
Qy 121 PALGRRPWLGRASSKASTPKPGTGPVPSFAEKVSDPEPPQLPEPQPRPRLQHLQASLS 180
Db 121 PALGRRPWLGRASSKASTPKPGTGPVPSFAEKVSDPEPPQLPEPQPRPRLQHLQASLS 180
Qy 181 QRLGSLDPGWLQCHSEVPDFLGAPKACRPDLGSEBSQLLIPGESAVLPGAGSQGPPEAS 240
Db 181 QRLGSLDPGWLQCHSEVPDFLGAPKACRPDLGSEBSQLLIPGESAVLPGAGSQGPPEAS 240
Qy 241 AFOEVSIRVSGSPQSSSGGKREWNPEEPWESPAQVQOESSQAGPPSPSEGAGAVAEEDPPG 300
Db 241 AFOEVSIRVSGSPQSSSGGKREWNPEEPWESPAQVQOESSQAGPPSPSEGAGAVAEEDPPG 300
Qy 301 EPVQAPPPQCCSPSPNRYHGLSPSSQARAGKAEGTAPLHIFPRLARHDRGNVYRLNMKQ 360
Db 301 EPVQAPPPQCCSPSPNRYHGLSPSSQARAGKAEGTAPLHIFPRLARHDRGNVYRLNMKQ 360
Qy 361 KHYVGRALRSRLRLKQAWKRWKKGCFGGGATVTTKSCFLNEQFQDHWAAQCRPPA 420
Db 361 KHYVGRALRSRLRLKQAWKRWKKGCFGGGATVTTKSCFLNEQFQDHWAAQCRPPA 420
Qy 421 SEEDTDAVGPEPLVPSPQVPEVPSLDPTVLPLSLGSPGQLAETPAEVFOALQOLGHOA 480
Db 421 SEEDTDAVGPEPLVPSPQVPEVPSLDPTVLPLSLGSPGQLAETPAEVFOALQOLGHOA 480
Qy 481 FRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVWSPLLSLMDDQ 540
Db 481 FRPQGERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRSPCLTLVWSPLLSLMDDQ 540
Qy 541 VSGLPCLKAACIHSWTRKQRESVLQKIRAAOVHVLMTPEALVAGGLPPAAQLPPVA 600
Db 541 VSGLPCLKAACIHSWTRKQRESVLQKIRAAOVHVLMTPEALVAGGLPPAAQLPPVA 600
Qy 601 FACIDEAHCLSQWSHNFRPCYLKVRCKLREMGVHCFGLGTATATRTASDVAAHVAE 660
Db 601 FACIDEAHCLSQWSHNFRPCYLKVRCKLREMGVHCFGLGTATATRTASDVAAHVAE 660
Qy 661 EPDLHGPAVPVPTNLHLVSMDRDTDQALLTLLOGKRFQNLDSIIYCNRRREDTERIAALL 720
Db 661 EPDLHGPAVPVPTNLHLVSMDRDTDQALLTLLOGKRFQNLDSIIYCNRRREDTERIAALL 720
Qy 721 RTCLHAAWPGSGGRAPKTTAEAYHAGMCSRERRVQRAFMQGLRVVVAATVAFGMGLDR 780
Db 721 RTCLHAAWPGSGGRAPKTTAEAYHAGMCSRERRVQRAFMQGLRVVVAATVAFGMGLDR 780
Qy 781 PDVRAVLHLGLPPSPFESYVQAVGRAGDQGAHCHLFLQPGEDLRELRRHVHADSTDFL 840
Db 781 PDVRAVLHLGLPPSPFESYVQAVGRAGDQGAHCHLFLQPGEDLRELRRHVHADSTDFL 840
Qy 841 AVKRLVQRVFPACTCTCTRPPSEQEGAVGGERPVKYPQAEQLSHQAAPGPRRVCMGH 900
Db 841 AVKRLVQRVFPACTCTCTRPPSEQEGAVGGERPVKYPQAEQLSHQAAPGPRRVCMGH 900

QY 901 ERALPIQLTVQALDMPERAIETLLCYLELHPHMLLELLATTYTHCRNLCPGPAQLOALA 960
 Db 901 ERALPIQLTVQALDMPERAIETLLCYLELHPHMLLELLATTYTHCRNLCPGPAQLOALA 960
 QY 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQWDHEPRTGV 1020
 Db 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQWDHEPRTGV 1020
 QY 1021 RRGTVLVEFSELAFLHRS PGDLTAEEKDQICDFLYGRVQARERQALRLRTTFOAFHSV 1080
 Db 1021 RRGTVLVEFSELAFLHRS PGDLTAEEKDQICDFLYGRVQARERQALRLRTTFOAFHSV 1080
 QY 1081 APPSCGCPLEQODEERSTRLLKDLLGRYFEEBQSPGGMEDAQGPQOARLQDWDQVR 1140
 Db 1081 APPSCGCPLEQODEERSTRLLKDLLGRYFEEBQSPGGMEDAQGPQOARLQDWDQVR 1140
 QY 1141 CDIROFLSLRPEEKFSRAVARIFHIGSPCYPAQVYGQDRFRWKYHLHLSPHALVGLAT 1200
 Db 1141 CDIROFLSLRPEEKFSRAVARIFHIGSPCYPAQVYGQDRFRWKYHLHLSPHALVGLAT 1200
 QY 1201 BELLQVAR 1208
 Db 1201 BELLQVAR 1208
 RESULT 3
 US-10-882-405-4
 ; Sequence 4, Application US/10882405
 ; Publication No. US20040259147A1
 ; GENERAL INFORMATION:
 ; APPLICANT: AGENE Research Institute, Co., Ltd.
 ; TITLE OF INVENTION: Gene causative of Rothmund-Thomson syndrome
 ; FILE REFERENCE: A1-003PCT
 ; CURRENT APPLICATION NUMBER: US/10/882,405
 ; CURRENT FILING DATE: 2004-07-02
 ; PRIOR APPLICATION NUMBER: US/09/889,325
 ; PRIOR FILING DATE: 2001-12-31
 ; PRIOR APPLICATION NUMBER: JP 1999-11218
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: Patent Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 1208
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-882-405-4
 Query Match 100.0%; Score 6424; DB 5; Length 1208;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MERLDRVRLQWERAAPRRGRRRSQDDVEAAPPETRALRYEYTLKRTTGOAGGGLR 60
 Db 1 MERLDRVRLQWERAAPRRGRRRSQDDVEAAPPETRALRYEYTLKRTTGOAGGGLR 60
 QY 61 SSESPLAAAEAEPEPCWPHLNRAATKSPQTPGRSROGSVDYQORLKANLKGTLQAG 120
 Db 61 SSESPLAAAEAEPEPCWPHLNRAATKSPQTPGRSROGSVDYQORLKANLKGTLQAG 120
 QY 121 PALGREPWLGRASSKASTPKPGTGVFVSFAEKVSDPEPQLPEPQPRGRQLQASLS 180
 Db 121 PALGREPWLGRASSKASTPKPGTGVFVSFAEKVSDPEPQLPEPQPRGRQLQASLS 180
 QY 181 QRLGSLDPGWLQCHSEVDFLQAPKACRPDLGSESSQLLIPGESAVLPGAGSQGPAS 240
 Db 181 QRLGSLDPGWLQCHSEVDFLQAPKACRPDLGSESSQLLIPGESAVLPGAGSQGPAS 240
 QY 241 AFOEVSIRVGSPOSSSGGKRWNEEPWESPAQVOQESSQAGPPSEGGAGAVAEEDPPG 300
 Db 241 AFOEVSIRVGSPOSSSGGKRWNEEPWESPAQVOQESSQAGPPSEGGAGAVAEEDPPG 300

QY 301 EPVQAQPPQCSPPSNRYHGLSPSSQARAKAGTAPLHIFPELARHDRGNVRLNWKQ 360
 Db 301 EPVQAQPPQCSPPSNRYHGLSPSSQARAKAGTAPLHIFPELARHDRGNVRLNWKQ 360
 QY 361 KHYVYRGALRSRLRKQAKWKKEGCEFGGGATVTTKESCFNLNQFQDHAAQCPRA 420
 Db 361 KHYVYRGALRSRLRKQAKWKKEGCEFGGGATVTTKESCFNLNQFQDHAAQCPRA 420
 QY 421 SEEDTAVGPEPLVPSQPPVEPSLPTVLPYLSIGSPGQLAETPAEVFQALQGHQA 480
 Db 421 SEEDTAVGPEPLVPSQPPVEPSLPTVLPYLSIGSPGQLAETPAEVFQALQGHQA 480
 QY 481 FRPQERAVWMLSGISTLLVLPYLSIGSPGQLAETPAEVFQALQGHQA 540
 Db 481 FRPQERAVWMLSGISTLLVLPYLSIGSPGQLAETPAEVFQALQGHQA 540
 QY 541 VSGLPPLCKAACIHSGMTKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAAQLPPVA 600
 Db 541 VSGLPPLCKAACIHSGMTKQRESVLQKIRAAQVHVLMLTPEALVGAGGLPPAAQLPPVA 600
 QY 601 PACIDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGLTATATRTTASDVQHLVAE 660
 Db 601 PACIDEAHCLSQSHNFRPCYLRVCKVLRERMGVHCFLGLTATATRTTASDVQHLVAE 660
 QY 661 EPDLHGPAVPNTLHLSVSMDDRTDQALLTLQGRFONLDSIIYCNRRDTERIAALL 720
 Db 661 EPDLHGPAVPNTLHLSVSMDDRTDQALLTLQGRFONLDSIIYCNRRDTERIAALL 720
 QY 721 RTCLHAAMVPGSGGRAPKTTAEAYHAGMCSRERRVQAFMQGQLRVVVVATVAFMGGLDR 780
 Db 721 RTCLHAAMVPGSGGRAPKTTAEAYHAGMCSRERRVQAFMQGQLRVVVVATVAFMGGLDR 780
 QY 781 PDVRAVLHGLPPSPFESTVQAVGRAGRDQPAHCHLFLQPGQEDLRELRRHVHADSTDFL 840
 Db 781 PDVRAVLHGLPPSPFESTVQAVGRAGRDQPAHCHLFLQPGQEDLRELRRHVHADSTDFL 840
 QY 841 AVKRLVQRFVFACTCTCTRPSEBQGVGGERPVKYPPOEAEQLSHQAAQPGPRVCMGH 900
 Db 841 AVKRLVQRFVFACTCTCTRPSEBQGVGGERPVKYPPOEAEQLSHQAAQPGPRVCMGH 900
 QY 901 ERALPIQLTVQALDMPERAIETLLCYLELHPHMLLELLATTYTHCRNLCPGPAQLOALA 960
 Db 901 ERALPIQLTVQALDMPERAIETLLCYLELHPHMLLELLATTYTHCRNLCPGPAQLOALA 960
 QY 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQWDHEPRTGV 1020
 Db 961 HRCPLAVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRALCOLQWDHEPRTGV 1020
 QY 1021 RRGTVLVEFSELAFLHRS PGDLTAEEKDQICDFLYGRVQARERQALRLRTTFOAFHSV 1080
 Db 1021 RRGTVLVEFSELAFLHRS PGDLTAEEKDQICDFLYGRVQARERQALRLRTTFOAFHSV 1080
 QY 1081 APPSCGCPLEQODEERSTRLLKDLLGRYFEEBQSPGGMEDAQGPQOARLQDWDQVR 1140
 Db 1081 APPSCGCPLEQODEERSTRLLKDLLGRYFEEBQSPGGMEDAQGPQOARLQDWDQVR 1140
 QY 1141 CDIROFLSLRPEEKFSRAVARIFHIGSPCYPAQVYGQDRFRWKYHLHLSPHALVGLAT 1200
 Db 1141 CDIROFLSLRPEEKFSRAVARIFHIGSPCYPAQVYGQDRFRWKYHLHLSPHALVGLAT 1200
 QY 1201 BELLQVAR 1208
 Db 1201 BELLQVAR 1208

RESULT 4

US-10-756-149-5768
 ; Sequence 5768, Application US/10756149
 ; Publication No. US20050181375A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Zlotnik, Albert
 ; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND

```
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756.149
; CURRENT FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5768
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5768

Query Match      100.0%; Score 6424; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MERLRDVRERLQAWERAFRRQRGRPPSQDDVEAAPEETRALYRYRTLKRTTGQAGGLR 60
Db 1 MERLRDVRERLQAWERAFRRQRGRPPSQDDVEAAPEETRALYRYRTLKRTTGQAGGLR 60
Qy 61 SSESIPAAAEAPRRCWGPPLHNRATKSPQTPGRSRQSGVDPYQORLKANLKGTLQAG 120
Db 61 SSESIPAAAEAPRRCWGPPLHNRATKSPQTPGRSRQSGVDPYQORLKANLKGTLQAG 120
Qy 121 PALGRRPWPLGRASKASTPKPGTGPVPSFAEKVSDPEPPQPPRPGRLQHLQASLS 180
Db 121 PALGRRPWPLGRASKASTPKPGTGPVPSFAEKVSDPEPPQPPRPGRLQHLQASLS 180
Qy 181 ORLGSLDPGWLORCHSEVDPFLGAPKACRPDLGSEESQLLIPGESAVILPGGAGSQGPAS 240
Db 181 ORLGSLDPGWLORCHSEVDPFLGAPKACRPDLGSEESQLLIPGESAVILPGGAGSQGPAS 240
Qy 241 AFQVSVIRVSGPQPSGSGEKRRWNEPWPESPAQVQOESSQAGPSPSEGAGAVAEEDPPG 300
Db 241 AFQVSVIRVSGPQPSGSGEKRRWNEPWPESPAQVQOESSQAGPSPSEGAGAVAEEDPPG 300
Qy 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAGKAGTAPLHIFPRLARHNRGNYVRLNMKQ 360
Db 301 EPVQAQPPQPCSSPNRYHGLSPSSQARAGKAGTAPLHIFPRLARHNRGNYVRLNMKQ 360
Qy 361 KHYVRGALRSRLRKQAWKQKKECFGGGATVTTKESCFNLNQDFHWAACQRPRA 420
Db 361 KHYVRGALRSRLRKQAWKQKKECFGGGATVTTKESCFNLNQDFHWAACQRPRA 420
Qy 421 SEEDTAVGPEPLVPSQPVEVPSLDPVTPLVLSLPGSGQLAETPAEVFOALQOLGHOA 480
Db 421 SEEDTAVGPEPLVPSQPVEVPSLDPVTPLVLSLPGSGQLAETPAEVFOALQOLGHOA 480
Qy 481 FRPQERAVNRILSGISTLLVLPFGAGKSLCYQLPALLYRRSPCLTLVWSPLLSLMDDQ 540
Db 481 FRPQERAVNRILSGISTLLVLPFGAGKSLCYQLPALLYRRSPCLTLVWSPLLSLMDDQ 540
Qy 541 VSGLPCCLKACIHSGMTKQRESVLQIRAAQVHVLMLTPEALVGAGGLPPAAQLPPVA 600
Db 541 VSGLPCCLKACIHSGMTKQRESVLQIRAAQVHVLMLTPEALVGAGGLPPAAQLPPVA 600
Qy 601 FACIDEAHCILSQWSHNRPFCYLRCKVLRERMGVHCFGLGTATATRTASDVAGHLAAVE 660
Db 601 FACIDEAHCILSQWSHNRPFCYLRCKVLRERMGVHCFGLGTATATRTASDVAGHLAAVE 660
Qy 661 EPDLHGPAVPTNLHLVSMRDRDQALLTLLOGKRFQNLDSIIYCNRRREDTERIAALL 720
Db 661 EPDLHGPAVPTNLHLVSMRDRDQALLTLLOGKRFQNLDSIIYCNRRREDTERIAALL 720
Qy 721 RTCLHAHVPGSGGRAPKTTAAETHAGCSERRRVOBAFMQGLRVVVAATVAFGMGLDR 780
Db 721 RTCLHAHVPGSGGRAPKTTAAETHAGCSERRRVOBAFMQGLRVVVAATVAFGMGLDR 780
Qy 781 PDVRAVLHLGLPPSFESVVOAGRAGRDGQPAHCHLFLQOGEDLRELRLRHVHADSTDFL 840
Db 781 PDVRAVLHLGLPPSFESVVOAGRAGRDGQPAHCHLFLQOGEDLRELRLRHVHADSTDFL 840
Qy 841 AVKRLVQRVFPACTCTCTTRPPSEOGAVGGERPVPKYPPOEAEQLSHQAAAPGPRRVCWGH 900
```

```
Db 841 AVKRLVQRVFPACTCTCTTRPPSEOGAVGGERPVPKYPPOEAEQLSHQAAAPGPRRVCWGH 900
Qy 901 ERALPIQLTVOALDMPPEAEITLLCYLELHPHWLELLATYTHCRINCPCGGPAQLQALA 960
Db 901 ERALPIQLTVOALDMPPEAEITLLCYLELHPHWLELLATYTHCRINCPCGGPAQLQALA 960
Qy 961 HRCPPPLAVCLAQQLPEDPGOGSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPTGV 1020
Db 961 HRCPPPLAVCLAQQLPEDPGOGSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPTGV 1020
Qy 1021 RRGTVLVFSELAFLHLSRSPDLTAEKQOICDPLYGRVQARERQALRLRRTFOAFHSV 1080
Db 1021 RRGTVLVFSELAFLHLSRSPDLTAEKQOICDPLYGRVQARERQALRLRRTFOAFHSV 1080
Qy 1081 AFPSCGPCLFQOQDEERSRLKDLILGRYFEEEGQEPGGMEDAQGPFGQARLQWEDQVR 1140
Db 1081 AFPSCGPCLFQOQDEERSRLKDLILGRYFEEEGQEPGGMEDAQGPFGQARLQWEDQVR 1140
Qy 1141 CDIRQFLSLRPEKFSRAVARIFPHGIGSPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT 1200
Db 1141 CDIRQFLSLRPEKFSRAVARIFPHGIGSPCYPAQVYGQDRRFWRKYLHLSFHALVGLAT 1200
Qy 1201 EELLQVAR 1208
Db 1201 EELLQVAR 1208

RESULT 5
US-11-084-955-2
; Sequence 2, Application US/11084955
; Publication No. US20050183149A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Radiological Sciences
; TITLE OF INVENTION: Targeted knockout mouse for human Rothmund-Thomson syndrome and
; FILE REFERENCE: preparation method thereof
; CURRENT APPLICATION NUMBER: US/11/084,955
; CURRENT FILING DATE: 2005-03-21
; PRIOR APPLICATION NUMBER: JP 2003-185409
; PRIOR FILING DATE: 2003-06-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1216
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-084-955-2

Query Match      61.4%; Score 3944; DB 6; Length 1216;
Best Local Similarity 63.9%; Pred. No. 1.2e-248;
Matches 795; Conservative 110; Mismatches 267; Indels 72; Gaps 13;

Qy 1 MERLRDVRERLQAWERAFRRQRGRPPSQDDVEAAPEETRALYRYRTLKRTTGQAGGLR 60
Db 1 MERLATVTRALQWERAFARLHGRPPAKGDVEAAPETRALYREYNLKVQVRQADRRH 60
Qy 61 SSE-SLPAAAEAPRRCWGPPLHNRATKSPQTPGRSRQSGVDPYQORLKANLKGTLQ 119
Db 61 VLEQSLAAEAEQEPSCWGPFLSRAATQNTQMPKQSLSSVQDYGKRLKANLKNNTQT 120
Qy 120 GPALGRRPWPLGRASKASTPKPGTGPVPSFAEKVSDPEPPQPPRPGRLQHLQASL 179
Db 121 GPTQSRKLQLOKRSLSLTVPAPEPPGSKTESPCDEADDDALPRVPEPRPGRLQHLQASL 180
Qy 180 SQRLGSLDPGWLORCHSEVDPFLGAPKACRPDLGSEESQLLIPGESAVILPGGAGS 235
Db 181 SRRLTSLDPGWLORCHSEVDPFLGAPKACRPDLGSEESQLLIPGESAVILPGGAGS 240
Qy 236 GPASAFQEVSVIRVSGPQPSGSGEKRRWNEPWPESPAQVQOESSQAGPSPSEGAGAVAE 295
Db 241 SPEAIAQQAQVLSQSPKSIKSKGRKKNK-K-GEDFAQ-DQPSGAGPLSEGARATVHG 298
```

```

Qy 296 EDPGCEPVQAPQPCSPSPNRYHGLSPSSQARAGKAGTAPLHIPPRLARHGRNVR 355
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 QDPFCEPTQNVNVPQPCNS-----SNQARTKAKGTHLHSPRPSASLDGRNVR 347
Qy 356 LNMOKKHVVRGALRSRLLRKQAWKQKKGCEGFGGGAATVTKESQFNEQDHWAAQ 415
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
348 LNMKNKRVVRVANGRLRLRKQVWKQKQKQAFGGSGPRATDKDTCFCGQGFHWASQ 407
Qy 416 CPRPA-----SEEDTDVAGP-----EPLVPSQPQVPEVPSL 446
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
408 CSQPGTLTVQBEGRDDKQPISTLEEVQARTGTASCHSGEETQPAAPELQ-VPHCPT- 465
Qy 447 DPTVLPLYSLSGSGQLAETPAEVQALQOLHQAAPRQOBRVAVRILSGISTLVLPTGA 506
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
466 --PMSPLYPPLGQVQAEPAEVQALRGLGRAPRQOERAIMRILSGISTLVLPTGA 523
Qy 507 GKSCLYQPLALLYSRRSCLTLVSPILSLMDQVSGPLPCLKAACIHSGMTKQRESVL 566
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 GKSCLYQPLALLYAQRSCLTLVSPILSLMDQVSDLPCLKAACIHSGMTKQRESVL 583
Qy 567 QKIRAAQVHVLMLTPEALVGA-----GGLPPAAQLPPVAFACIDBAHCLSQSHNFRPCY 621
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
584 KKVRAAQVHVLIVSPEALVGCARGPSLPOAAQLPPIAFACIDBHVHCLSQSHNFRPCY 643
Qy 622 LRVCVLRERMGVHCFGLGTATARTASDVAAHLAVAEEDPLHGPAPVPTNLHLSVMD 681
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 LRVCVLRERMGVHCFGLGTATARTASDVAAHLAVAEEDPLHGPAPVPTNLHLSVMD 703
Qy 682 RDTQALLTLQGRFONLDSIIICNRRDTERIAALLRTCLHAAWVPGSGRAPKTTA 741
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
704 RDSQALLVTLQGRFRLDSVIICTRIERQNGWALLRLCLSVGDSRRGCGPEAIA 763
Qy 742 EAYHAGMCSRERRRVRAPMOCQLAVVATVAFMGMLDRPDVRAVLHLGLPPSFESYVQA 801
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
764 EAYHAGMSSQERRRVRQAFMRGHLRMVATVAFMGMLDRPDVRAVLHLGLPPSFESYVQA 823
Qy 802 VBRAGDQOPACHLFLQOGEDILBELRHVHADSTDFLAVKRLVORVPEACTCTCTRP 861
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
824 IGRAGRDGKPAHCHLFMHPQGEDLWELRRHAHDSTDFLAVKRLVORVPEACTCTCT 878
Qy 862 SRQEGAVGERPVKVPQPEABOLLS-HQAAPGRVRCVCHGERALPIQLTVQALDMPBEAI 920
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
879 -----QRPVKSSEEVKEHSGOQTYFVLGOALCIGHERALPVQSTVQALDMPBEAI 929
Qy 921 ETLCLYLELPHHLLLEATTYTHCLNCPGQPAQLQALAHRCPPELAVCLAQOLPEDPCQ 980
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 ETLCLYLELPHHLLLELPWTYAQCHLHCLGSAQLQALAHRCPPELAAQAKWPKDTSQ 989
Qy 981 GSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRGTGVLVEFSELAFLHLSR 1040
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 GRSSLEFGVVELADSMGWKLASVROALHQLKWDPEPKGAAQGTGLVKFSELAFLHLSR 1049
Qy 1041 GDLTAREKDOIICDLYGRVQAREQALRLRTFOAHSVAFPSGPCLEQOEDBERSTRL 1100
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1050 GDLTAREKDOIICDLYGRVQAREQALRLRTFOAHSVAFPSGPCLEQOEDBERSTRL 1109
Qy 1101 KDLGRYFEEBQPGGMEQAGPEPQOARLQDWEQVRCDIROFLSLRPEKFSRRAV 1160
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1110 KTLVSIFYEEB-EBEETWTDTQGPQGTQLQDWEQVRCDIROFLSLRPEKFSRRAV 1168
Qy 1161 ARIFHIGSPCYPAQVYGDPRFWRKYLHLSPHALVGLATEELL 1204
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1169 ARIFHIGSPCYPAQVYGLDRFRWRKYLHLSPHALVGLATEELL 1212

```

RESULT 6

US-11-097-143-16743
; Sequence 16743, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

```

; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16743
; LENGTH: 1579
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-16743

```

```

Query Match 26.8%; Score 1723.5; DB 6; Length 1579;
Best Local Similarity 28.9%; Pred. No. 2.1e-103;
Matches 467; Conservative 217; Mismatches 459; Indels 471; Gaps 45;

Qy 10 RLQAWERAFRRQRORRPSQDDVEAAPBETRALYREYTLKKT----- 51
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
16 RVKWEKDFKKKNGRVPSKYDIRDASQEIIRDSYKMYKLTSTFLEETLNDVLSDEGYDIL 75
Qy 52 -TQAGG-----GLRS-----ESLP----- 66
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 EMSQASDFGVSMDDVSLNEGQPLDLSALVQSSGNLEEIPQSVESGFSNLIDLPLN 135
Qy 67 -----AAAEAP-EPRCWGPHLNAATKSPQ-----TPGRSRQ 100
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 RQVLTNLVRDENHVIRKFEAVEELPINQNAWGLNVKPPAPPQPVASKSAFGHGKQ- 194
Qy 101 SVDPYQRLKANLKTGL-QAGPALRR--PWPLGRASSKASTPKPGTGPVPSFAKVSD 157
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 --PRAGASLKPSSLAKLFPQSSRGFAKNPRKPLSCVSSSTTSLSSVPTDHHLELDF 252
Qy 158 EPPQLPEPQPRGRLO-----HLQASLQRLGSLDPGNLOR-----CHSEVP- 199
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 ETILIRKAEYKKEKQQAIAANNPLASEHSKESIKT---LVDDGMLRRNTKENTLDEEVPF 309
Qy 200 ---DFLGAPK-----ACRPDLGSEESQLLIPGESAVL-----GP 230
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 ABANNSTGSKTTFNGLANLDSLKLPVKEKVLQAKPDQWAIQELQTDNWSNMNQK 369
Qy 231 GAGSGPEASAFQEVIRVSGFPSS-----SGGEKRRWNEE----- 267
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 DHLNHTTPASSQKSVAPK-NKPPFPEQETDSDSVVAESEEEQEPQVYRQLSKRKIVS 428
Qy 268 -----PWESPAQVOQES---SQAGP----- 284
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
429 TASGKVEVAAPVEIPNKEPETETFAQENPDFSAEDQDATVYPENKKKAKRKAQAGK 488
Qy 285 -----PSEGAGAVAVEEDPPGPVQAOQPPQ--- 310
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
489 QKTTKPAEPKPKTEKKAQVAKKPAKPKPRNSKKAIAVEPADPDEDERQPLNPEDL 548
Qy 311 -----CSSPSNPRYH-----GLSPSSQARAGKAGTAP 338
Dy :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 KYVLALEAGDITSVPRINVQDLEADATAQRYIRTFRTAAGPNPGLSEGSNIRVDEKRAAR 608

```



```

Db 498 FVDMRSIIYCKFQAEITDFVSKYL---CDN-----NITAKSVHSGLLKNRSRV 543
QY 757 QRAFMQGLRVVAVTAFVGMGLDRPDRVAVLHGLPPSFESYQVAVGRAGRGOPAHCHL 816
Db 544 QELFCSNKIRVP-----SY-----NF 559
QY 817 FLOPQEDLRLRHVHADSTDFLAVKLVQVRVPACTCTTRPPSEOGVAGGERVPVK 876
Db 560 TEQP-----FLGSDGVDGYAMSKFLYQIF-----SSENTTG-----590
QY 877 YPQBAEQLSHQAPRRVCMGHERALPIQLTVQALDMPBEAIEITLLCYLELHPHWLE 936
Db 591 -----CIG-----SLAKELTSKFDIKEVLTILTQLEIGDQOYIR 627
QY 937 LIATYTHCRNCPGPAQLQALAHRCPPPLAVC-----LAQQLPEDPGQSSSVEFDVVKL 992
Db 628 LLPQFSVTCVL-----YFKTSPQLLADKDLILRSVLRNRMKDGHYVFDIPRI 676
QY 993 VDSMGWELASVRRALCQOWDHEPTGVRGTGVLVFES-ELAFHLRSPG-----1041
Db 677 ANDLKITMNEV-----FDHLHK-----LKFSGEISFELKDPAYCYVILWRPD 718
QY 1042 DLTASEKQICDFYGRVQARERQALRLRRTFQAFHSVAPPSC-----GPCLE 1090
Db 719 DFNALSAN-----LTKWISEVSSKISKLDAMP-ALANPAVKGCKRTGCGSSQHTPCIQ 772
QY 1091 QODBERSTRKDLGLGRYFEEBEGPFGMEDAQGPPEQOARLODWEQVRCDIROFLSLR 1150
Db 773 KKIME-----YFSKODGTSEND-RTQLQKSPFLQADIKVFIQSN 812
QY 1151 PREKSSRAVARIFIGISPCYPAQYQGDPRFWRKYLHLSHALVGLATELLOV 1206
Db 813 SPAKTPRAVARIMEGISPAPSPVTSKN-HFWGRYVEVDFFLYMAAKAELVKL 867

```

RESULT 8

```

US-10-374-077-75
; Sequence 75, Application US/10374077
; Publication No. US20040006779A1
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; Oshima, Junko
; Mulligan, John T.
; Schellenberg, Gerald D.
; TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO
; WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/374,077
; FILING DATE: 25-Feb-2003
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman, Stephen
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 100107.401D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 607 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-10-374-077-75

Query Match Similarity 9.8%; Score 632.5; DB 4; Length 607;
Best Local Similarity 37.6%; Pred. No. 1.2e-32;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

QY 474 EQLGQAFRQPRQAVRNVRIISGISTLVLPFGAGKSLCYQLPALLYSRSPCLTIVVSPL 533
Db 18 ETFGYQFRPQREIITVLGRDCLVVMPTGGKSLCYQIPALLNG----LTVVVSPL 73
QY 534 LSLMDDQVSGLPP-CLKAACTHSGMTRKQRESVLQKRAAQVHVLMLTPEALVAGAGGLPP 592
Db 74 ISLMKDQVDQLQANGVAAACLNSTQTRQQLLEVMTGCTGQIRLLYIAPERLMDNFLEH 133
QY 593 AAQLPPVAFACIDEAHCLISQWSHNRPCYLKRVCKVLBERMGVHCFGLGTATATRTASDV 652
Db 134 LAHWNPVLLA-VDEAHCIQSGHDFRPEYALGQ-LRQRFPTLPPMALATATADDTTQDI 191
QY 653 AQLHVAEEDPLHGPAPVPTNLHLSVSMDDRTDQALLTLQCKRFONLDSI-----703
Db 192 VRLG-----LNDPL-----IQIS-SFDRNIRYMLM-----EKEKPLDQLMRYVQEQRG 235
QY 704 ---IYCNRRDTERIAALLRTCLHAAWVPGSGRPAKTTABAYHAGMCSRERRRVQRAF 760
Db 236 KSGIYCNRAKVEDTAAL-----QSKGISAAAVHAGLENNVRADVQEK 281
QY 761 MQQLRVVTVAVFGMLDRPDVRAVLHGLPPSFESYQVAVGRAGRGOPAHCHLFL-- 818
Db 282 QRDDLQIVVATVAFGMGINKENVRVVFHFDIPRNIESYQETGRAGRDGLPAEAMLFYDP 341
QY 819 -----OPQGEDLRELRRH---VHADSTDTFLAVKRLV-----QRVFPACTC- 855
Db 342 ADMAWLRRCLEKFGQ-LQDIERHKLNAMGAFABEAQTCLRLVLLNYFGEGRQEQCGNCD 400
QY 856 TCTRPPSEQEGAVGGE 871
Db 401 ICLDPPKQYDGDSTDAQ 416

```

RESULT 9

```

US-09-815-242-10414
; Sequence 10414, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```

```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10414
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10414

Query Match
Best Local Similarity 9.8%; Score 631.5; DB 3; Length 610;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

Qy 474 EQLGHQAFRPGQBRVAVNRILSGISTLVLPTGAGKSLCYQLPALLYRRSPCLTLVVSPL 533
Db 21 ETFGYQFRPGQBEIITVLGRDCLVMPPTGGKSLCYQIPALLNG----LTVVVSPL 76
Qy 534 LSLMDDQVSGLPP-CLKAACIHSGMTKQRESVLQKIRAAQVHVLMLTPEALVAGAGLPP 592
Db 77 ISLMKDQVDQLQANGVAAACLNSTQREQLLEVMTGCTGQIRLLYIAPERLMLDNFLEH 136
Qy 593 AAOPLPVAFACIDEAHCLQSOWSHNFRPCYLKRVCKLBERMGVHCFLGLTATARTASDV 652
Db 137 LAHNVPVLLA-VDEAHCSISQWGHDFRPEYALGQ-LRQRFPTLPFMALTATADTTTQDI 194
Qy 653 AQHLAVAEEDPLHGPAPVPTNLHLSVSMRDRTDQALLTLQGRFQNLDSI----- 703
Db 195 VRLIG-----LNDEL-----IQIS-SFDRPNIRYMLM-----EKFPLDQLMRVYQSQRG 238
Qy 704 ---IYCNRRDTERIAALLTCLHAAWPGSGRAPKTTABAYHAGMCSRERRVQRAF 760
Db 239 KSGIIVCNSRAKVEDTAARL-----QSKGISAAAYHAGLENNVRADVQEKF 284
Qy 761 MQQLRVVATVAFGMGLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGPACHLFL-- 818
Db 285 QRDDLQIVVATVAFGMGINKPNRVFVHFDIPRNIESYYQETGRAGRDGLPAEAMLYDP 344
Qy 819 -----OPQGEDLRELRRH---VHADSTDFLAVKRLV-----QRVPACTC- 855
Db 345 ADMAWLRCLLEKEFGQO-LQDIERHKLNAMGAPAEQTCRRLVLLNYFGEGRQPCGNC 403
Qy 856 TCTRPPSEQEAGVGGE 871
Db 404 ICLDPPKQYDGSSTDAQ 419
```

```
RESULT 10
US-10-282-122A-43357
; Sequence 43357, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43357
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43357

Query Match
Best Local Similarity 9.8%; Score 631.5; DB 4; Length 610;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;

Qy 474 EQLGHQAFRPGQBRVAVNRILSGISTLVLPTGAGKSLCYQLPALLYRRSPCLTLVVSPL 533
Db 21 ETFGYQFRPGQBEIITVLGRDCLVMPPTGGKSLCYQIPALLNG----LTVVVSPL 76
Qy 534 LSLMDDQVSGLPP-CLKAACIHSGMTKQRESVLQKIRAAQVHVLMLTPEALVAGAGLPP 592
Db 77 ISLMKDQVDQLQANGVAAACLNSTQREQLLEVMTGCTGQIRLLYIAPERLMLDNFLEH 136
Qy 593 AAOPLPVAFACIDEAHCLQSOWSHNFRPCYLKRVCKLBERMGVHCFLGLTATARTASDV 652
Db 137 LAHNVPVLLA-VDEAHCSISQWGHDFRPEYALGQ-LRQRFPTLPFMALTATADTTTQDI 194
Qy 653 AQHLAVAEEDPLHGPAPVPTNLHLSVSMRDRTDQALLTLQGRFQNLDSI----- 703
Db 195 VRLIG-----LNDEL-----IQIS-SFDRPNIRYMLM-----EKFPLDQLMRVYQSQRG 238
Qy 704 ---IYCNRRDTERIAALLTCLHAAWPGSGRAPKTTABAYHAGMCSRERRVQRAF 760
Db 239 KSGIIVCNSRAKVEDTAARL-----QSKGISAAAYHAGLENNVRADVQEKF 284
Qy 761 MQQLRVVATVAFGMGLDRPDVRAVLHLGLPPSFESYVQAVGRAGRDGPACHLFL-- 818
Db 285 QRDDLQIVVATVAFGMGINKPNRVFVHFDIPRNIESYYQETGRAGRDGLPAEAMLYDP 344
Qy 819 -----OPQGEDLRELRRH---VHADSTDFLAVKRLV-----QRVPACTC- 855
Db 345 ADMAWLRCLLEKEFGQO-LQDIERHKLNAMGAPAEQTCRRLVLLNYFGEGRQPCGNC 403
Qy 856 TCTRPPSEQEAGVGGE 871
Db 404 ICLDPPKQYDGSSTDAQ 419
```

```
RESULT 11
US-09-753-143-77
; Sequence 77, Application US/09753143
; Patent No. US20020102550A1
; GENERAL INFORMATION:
; APPLICANT: NATHAN A. ELLIS, JAMES GERMAN, AND JOANNA
; GRODEN
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
```

—
—
**
**
**
—
**
—
—
—
**

Matches	162;	Conservative	58;	Mismatches	137;	Indels	73;	Gaps	14;
QY	474	EQLGHQAFRGQBRAMRILISGISTLLVLP	TGAGKSLCYQLPALLYRRRSPCLTLVVSPL	533					
Db	25	ETFGYQQFRGQBAI	IDTALSGRDCLVMP	TGGGKSLCYQIPALLDQ---	LTVVVSPL	80			
QY	534	LSLWMDQVSG-LPPCLAACTHSGMTKQRESVLQ	KRAAQVHVLMLTPEALVGAGGLPP	592					
Db	81	ISLWKQQVDQLLANGVAACLNSTQSRQQQL	EVNAGCTGQIRILYTAPERLMLDNFLDH	140					
QY	593	AAQLPPVAFACIDEAHCLSQWSHNFRPCYLRV	CKVLRRMGVHCFGLGTATATRTTASDV	652					
Db	141	LAHWNPVLLA-VDEAHCTISQWGHDFR	PEYAAALQ-LRQRFALPFMAL	TATADDTTRQDI	198				
QY	653	AQHILAAVEPDLHGCPAPVP	PNLHLVSMVDRDQDQALLTLLOGKEFONLDSI-	703					
Db	199	IRLLG-----LNDPL-----	IQIS-SFDRPNIRYMLM-----	EXFKPJDQLMRYVQEQRG	242				
QY	704	---IITYCNREDTERIAALLRTCLHAAV	FGSGGRAPKTTAEAYHAGMCSRERRRVQRAF	760					
Db	243	KSGLIYCNRAKVEDTAARLQ-----	SRG-----ISAAVHAGLENAIRADVQKF	288					
QY	761	MQQGLRVVVTATVAFGMGLDRPDVRAVLH	GLHPSPSFESVQAVGRGRDQGAHCHLFLQP	820					
Db	289	QRDDLQIVVATVAFGMGINPENRVFVHF	DFPRNIESYQETGRAGRDGLPAEAMLFYDP	348					
QY	821	-----QGEDLRERLRHHVHADSTDF	LAVKRLV-----QRVFPACTC-T	856					
Db	349	ADMANWLRCLEEKPAQQLQDIERHKLN	AMGAFAEAAQTCTRLVLNLYFGEGQEP	CGNCDI	408				
QY	857	CTRPPSQEQG	866						
Db	409	CLDPKQYDQ	418						

RESULT 15

US-10-282-122A-74973
; Sequence 74973, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:38:41 ; Search time 54.2 Seconds
(without alignments)
2448.199 Million cell updates/sec

Title: US-09-889-325-4_COPY_907_1208
Perfect score: 1617
Sequence: 1 QLTVALDMPBEAETLLCY.....HLSFHALVGLATELLQVAR 302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*
9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1617	100.0	1142	8	Abm84821 Human dia
2	1617	100.0	1144	7	Add01202 Human nuc
3	1617	100.0	1208	2	Aaw95050 Human hel
4	1617	100.0	1208	3	Aab20993 Human Rec
5	1617	100.0	1208	5	Abg93387 Human Rec
6	1617	100.0	1208	8	Adq21562 Human sof
7	1617	100.0	1208	8	Adu80701 Human RYS
8	1594.5	98.6	1306	8	Adm90956 Human pha
9	1144.5	70.8	1216	9	Adw44021 Mouse REC
10	427	26.4	1579	4	Abb63317 Drosophil
11	145.5	9.0	206	8	Ady13119 Plant ful
12	108.5	6.7	762	7	Abc80677 Pseudomon
13	107	6.6	751	6	AbR82253 Human nuc
14	107	6.6	755	8	Abm80914 Tumour-as
15	107	6.6	1029	3	Aay96736 PRO3434,
16	107	6.6	1029	3	Aab24027 Human PRO
17	107	6.6	1029	5	Abg92710 Human sec
18	107	6.6	1029	5	Aau77765 Tumour as
19	107	6.6	1029	5	Abg91362 Novel hum
20	107	6.6	1029	5	Abg31402 Human PRO
21	107	6.6	1029	6	Abu72378 Novel hum
22	107	6.6	1029	6	Abu80868 Human sec
23	107	6.6	1029	6	Abg73315 Human PRO
24	107	6.6	1029	6	Abu60814 Human sec

25	107	6.6	1029	6	ABU81237 Human PRO
26	107	6.6	1029	6	ABU62958 Human PRO
27	107	6.6	1029	7	ABO01895 Novel hum
28	107	6.6	1029	7	ADB89026 Human tum
29	107	6.6	1029	7	AAE38827 Human PRO
30	107	6.6	1029	7	ADJ58520 Human PRO
31	107	6.6	1029	8	ADI10343 Human PRO
32	107	6.6	1029	9	AEb47258 Human tum
33	107	6.6	1029	9	AEb70219 Human tum
34	107	6.6	1377	6	ABP56518 Human BAA
35	107	6.6	2190	7	ADB80326 Human MDD
36	106	6.6	253	8	ADQ66277 Novel hum
37	105.5	6.5	699	7	ADI21211 Novel hum
38	102.5	6.3	279	7	ABO70187 Pseudomon
39	102	6.3	817	7	ABO77141 Pseudomon
40	100	6.2	552	4	AAE01117 Human gen
41	100	6.2	552	5	ABg64586 Human alb
42	100	6.2	552	8	ADL77853 Albumin f
43	99.5	6.2	512	7	ABO73593 Pseudomon
44	98	6.1	719	4	AAM00798 Human bon
45	98	6.1	724	4	AAM00911 Human bon

ALIGNMENTS

RESULT 1
ABM84821
ID ABM84821 standard; protein; 1142 AA.
XX
AC ABM84821;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:5070.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patory S, Shi X, Suarez CJ;
XX
DR WPI: 2004-329368/30.
DR N-PSDB; ACN43473.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human

CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX

SQ Sequence 1142 AA;

Query Match 100.0%; Score 1617; DB 8; Length 1142;
Best Local Similarity 100.0%; Pred. No. 1.1e-158;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 QLTVAQLDMPPEAEITLLCYLELPHHLELLATTYTHCRLNCPGPAQALAHRCPPPL 60
Db 841 QLTVAQLDMPPEAEITLLCYLELPHHLELLATTYTHCRLNCPGPAQALAHRCPPPL 900
Qy 61 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRGTGV 120
Db 901 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRGTGV 960
Qy 121 LVFESELAHLRSPGDLTAEEKDQICDPLYGRVQARERQALRLRRTFOAFHSAVPSCG 180
Db 961 LVFESELAHLRSPGDLTAEEKDQICDPLYGRVQARERQALRLRRTFOAFHSAVPSCG 1020
Qy 181 PCLEQQDEERSTRKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIROF 240
Db 1021 PCLEQQDEERSTRKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIROF 1080
Qy 241 LSLRPEKFSRAVARIFHIGSPCYPAQVYGQDRFRWRKYLHLSFHALVGLATEELLQV 300
Db 1081 LSLRPEKFSRAVARIFHIGSPCYPAQVYGQDRFRWRKYLHLSFHALVGLATEELLQV 1140
Qy 301 AR 302
Db 1141 AR 1144

RESULT 2
ADD01202
ID ADD01202 standard; protein; 1144 AA.

XX ADD01202;
XX 01-JAN-2004 (first entry)

DE Human nucleic acid-associated protein NAAP-40 SEQ ID NO:40.

KW human; nucleic acid-associated protein; NAAP; cytostatic;
KW antiarteriosclerotic; anti-HIV; antiallergic; cerebroprotective;
KW antiparkinsonian; anticonvulsant; neurotropic; neuroprotective;
KW antiinflammatory; ophthalmological; thyromimetic; antiarthritis;
KW hepatocytic; antibacterial; virucide; protozoacide; antiparasitic;
KW fungicide; gene therapy; cell proliferative disease; cancer;
KW atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
KW Alzheimer's disease; stroke; epilepsy; developmental disorder;
KW renal tubular acidosis; anaemia; glaucoma; hypothyroidism;
KW autoimmune disorder; inflammatory disorder; AIDS; allergy;
KW atopic dermatitis; arthritis; infection.

OS Homo sapiens.
XX WO2003054219-A2.
XX 03-JUL-2003.

PF 18-DEC-2002; 2002MO-US041115.
XX
PR 19-DEC-2001; 2001US-0343004P.
PR 11-JAN-2002; 2002US-0347633P.
PR 25-JAN-2002; 2002US-0351749P.
PR 22-FEB-2002; 2002US-0359498P.
XX
(INCY-) INCYTE GENOMICS INC.
XX
PI Azimzai Y, Baughn MR, Becha SD, Borowsky ML, Burford N;
PI Elliott VS, Emerling BM, Forsythe IO, Gorvad AE, Griffin JA;
PI Kable AE, Khare R, Lal PG, Lee EA, Lee SY, Li JX, Marquis JP;
PI Ramkumar J, Richardson TW, Sprague WW, Swarnakar A, Tang VT;
PI Chawla NK, Warren BA, Yue H;
XX
DR WPI; 2003-559157/52.
DR N-PSDB; ADD01259.

XX New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
PT infections.

PS Claim 1; SEQ ID NO 40; 405pp; English.

XX The present invention describes human nucleic acid-associated proteins
CC designated NAAP-1 to NAAP-57. The human NAAPs have cytostatic,
CC antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
CC antiparkinsonian, anticonvulsant, neurotropic, neuroprotective,
CC antiinflammatory, ophthalmological, thyromimetic, antiarthritis,
CC hepatocytic, antibacterial, virucide, protozoacide, antiparasitic and
CC fungicide activities, and can be used in gene therapy. The NAAP protein
CC and polynucleotide sequences can be used in diagnosing, treating and
CC preventing diseases or conditions associated with the decreased
CC expression or overexpression of NAAP, such as cell proliferative diseases
CC (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
CC (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
CC developmental disorders (renal tubular acidosis, anaemia, glaucoma,
CC hypothyroidism), autoimmune/inflammatory disorders (AIDS, allergies,
CC atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
CC parasitic, protozoal, fungal). The present sequence represents human NAAP
XX -40, from the present invention.

SQ Sequence 1144 AA;

Query Match 100.0%; Score 1617; DB 7; Length 1144;
Best Local Similarity 100.0%; Pred. No. 1.1e-158;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLTVAQLDMPPEAEITLLCYLELPHHLELLATTYTHCRLNCPGPAQALAHRCPPPL 60
Db 843 QLTVAQLDMPPEAEITLLCYLELPHHLELLATTYTHCRLNCPGPAQALAHRCPPPL 902
Qy 61 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRGTGV 120
Db 903 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRGTGV 962
Qy 121 LVFESELAHLRSPGDLTAEEKDQICDPLYGRVQARERQALRLRRTFOAFHSAVPSCG 180
Db 963 LVFESELAHLRSPGDLTAEEKDQICDPLYGRVQARERQALRLRRTFOAFHSAVPSCG 1022
Qy 181 PCLEQQDEERSTRKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIROF 240
Db 1023 PCLEQQDEERSTRKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIROF 1082
Qy 241 LSLRPEKFSRAVARIFHIGSPCYPAQVYGQDRFRWRKYLHLSFHALVGLATEELLQV 300
Db 1083 LSLRPEKFSRAVARIFHIGSPCYPAQVYGQDRFRWRKYLHLSFHALVGLATEELLQV 1142
Qy 301 AR 302
Db 1143 AR 1144

RESULT 3
AAW95050
ID AAW95050 standard; protein; 1208 AA.

XX AC AAW95050;
XX DT 14-MAY-1999 (first entry)
XX DE Human helicase protein.

XX KW RecQ4 gene; helicase; Werner's syndrome; Bloom's syndrome; human.

XX OS Homo sapiens.
XX PN WO9905284-A1.
XX PD 04-FEB-1999.

XX PF 10-JUL-1998; 98WO-JP003114.
XX PR 25-JUL-1997; 97JP-00200387.

XX PA (AGEN-) AGENE RES INST CO LTD.
XX PI Shimamoto A, Kitao S, Furuichi Y;
XX DR WPI; 1999-142939/12.
XX DR N-PSDB; AAX21656.

XX PT New human helicase gene RecQ4 - used for investigation and diagnosis of
XX PT helicase-implicated diseases such as Werner's syndrome.
XX PS Claim 1; Page 35-42; 67pp; Japanese.

XX CC The present sequence represents a protein having helicase activity
XX CC encoded by the human gene RecQ4. The gene has significant homology to the
XX CC Escherichia coli helicase gene (RecQ). Host cells transformed with
XX CC vectors comprising the RecQ4 gene are used for the recombinant expression
XX CC of the protein. The gene may be used for the study and diagnosis of
XX CC disorders in which helicase activity is involved, such as Werner's and
XX CC Bloom's syndromes in which mutations in the helicase gene are implicated
XX SQ Sequence 1208 AA;

Query Match 100.0%; Score 1617; DB 2; Length 1208;
Best Local Similarity 100.0%; Pred. No. 1.2e-158;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVALDMPPEAETLLCYLELHPHMLLELLATTYTHCRINCPCGGPAQLQALAHRCPPPL 60
DB 907 QLTVALDMPPEAETLLCYLELHPHMLLELLATTYTHCRINCPCGGPAQLQALAHRCPPPL 966

QY 61 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRGTGTV 120
DB 967 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRGTGTV 1026

QY 121 LVFESELAFLHRS PGDLTAEKDKQICDPLYGRVQARERQALRLRTTFOAFHSVAFPPSCG 180
DB 1027 LVFESELAFLHRS PGDLTAEKDKQICDPLYGRVQARERQALRLRTTFOAFHSVAFPPSCG 1086

QY 181 PCLEQDDEERSTRLKDLGRYFEFEEBGPQGMEDAQGPQCARLQWEDQVRCDIIRQF 240
DB 1087 PCLEQDDEERSTRLKDLGRYFEFEEBGPQGMEDAQGPQCARLQWEDQVRCDIIRQF 1146

QY 241 LSLRPEEKSSAVARIIFHGIGSPCYPAQVYQDQRRFRWKYHLHSHALVGLATELLOV 300
DB 1147 LSLRPEEKSSAVARIIFHGIGSPCYPAQVYQDQRRFRWKYHLHSHALVGLATELLOV 1206

QY 301 AR 302
DB 1207 AR 1208

RESULT 4
AAB20993

XX ID AAB20993 standard; protein; 1208 AA.
XX AC AAB20993;

XX DT 11-DEC-2000 (first entry)

XX DE Human RecQ4 helicase.

XX KW RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;

XX KW poikiloderma congenitale; autosomal recessive; skin disorder;

XX KW dermatology; antibody; prenatal diagnosis; gene therapy.

XX OS Homo sapiens.
XX PN WO200043522-A1.
XX PD 27-JUL-2000.

XX PF 19-JAN-2000; 2000WO-JP000233.
XX PR 19-JAN-1999; 99JP-00011218.

XX PA (AGEN-) AGENE RES INST CO LTD.
XX PI Kitao S, Shimamoto A, Furuichi Y;
XX DR WPI; 2000-524241/47.
XX DR N-PSDB; AAA72320, AAA72321.

XX PT RecQ4 helicase gene, gene products and antibody, used in the diagnosis
XX PT and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.
XX PS Example 5; Page 83-92; 115pp; Japanese.

XX CC The present sequence represents human RecQ4 helicase. The invention
XX CC relates to the genomic DNA sequence of human RecQ4 helicase (AAA72320).
XX CC Mutations in the RecQ4 helicase gene, located on chromosome 8q24.3, are
XX CC the cause of Rothmund-Thomson syndrome (also known as poikiloderma
XX CC congenitale), an autosomal recessive skin disorder principally occurring
XX CC in females and often accompanied by juvenile cataracts, saddle nose,
XX CC congenital bone defects, hypogonadism and disturbances in the growth of
XX CC hair, nails and teeth. The invention also relates to vectors and host
XX CC cells comprising the human RecQ4 helicase genomic sequence. It
XX CC additionally encompasses use of the RecQ4 helicase protein as a
XX CC therapeutic and anti-RecQ4 antibodies as diagnostic agents. The RecQ4
XX CC helicase gene and its products, and anti-RecQ4 helicase antibodies are
XX CC useful in the diagnosis, especially prenatal diagnosis, and treatment of
XX CC Rothmund-Thomson syndrome. The genomic sequence may especially be used in
XX CC gene therapy for this condition

XX SQ Sequence 1208 AA;

Query Match 100.0%; Score 1617; DB 3; Length 1208;
Best Local Similarity 100.0%; Pred. No. 1.2e-158;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVALDMPPEAETLLCYLELHPHMLLELLATTYTHCRINCPCGGPAQLQALAHRCPPPL 60
DB 907 QLTVALDMPPEAETLLCYLELHPHMLLELLATTYTHCRINCPCGGPAQLQALAHRCPPPL 966

QY 61 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRGTGTV 120
DB 967 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRGTGTV 1026

QY 121 LVFESELAFLHRS PGDLTAEKDKQICDPLYGRVQARERQALRLRTTFOAFHSVAFPPSCG 180
DB 1027 LVFESELAFLHRS PGDLTAEKDKQICDPLYGRVQARERQALRLRTTFOAFHSVAFPPSCG 1086

QY 181 PCLEQDDEERSTRLKDLGRYFEFEEBGPQGMEDAQGPQCARLQWEDQVRCDIIRQF 240
DB 1207 AR 1208

Db 1087 PCLEQDEERSTRKDLGRYFEEEGQPGMEDAQGPQARLQDWDQVRCDIRQF 1146
QY 241 LSLRPEKFSRAVARIFHGIGSPCYPAQVYGQDRFRWKYHLHSFHALVGLATEELLQV 300
Db 1147 LSLRPEKFSRAVARIFHGIGSPCYPAQVYGQDRFRWKYHLHSFHALVGLATEELLQV 1206
QY 301 AR 302
Db 1207 AR 1208

RESULT 5
ABG93387
ID ABG93387 standard; protein; 1208 AA.
XX AC ABG93387;
XX 20-NOV-2002 (first entry)
XX Human RecQ protein-like 4 (RECQL4).
DE Human; RecQ protein-like 4; RECQL4; infection; inflammation; cytostatic;
KW tumour formation; cancer; antiinflammatory; antimicrobial;
KW antisenese therapy.
XX
OS Homo sapiens.
XX
XX US6436706-B1.
XX 20-AUG-2002.
XX 23-FEB-2001; 2001US-00792594.
XX 23-FEB-2001; 2001US-00792594.
XX (ISIS-) ISIS PHARM INC.
XX Ward DT, Watt AT;
XX WPI; 2002-689941/74.
XX N-PSDB; ABS68851.
XX
XX New antisense compounds targeted to nucleic acids encoding RecQ protein-like 4, useful for modulating expression of the nucleic acid and treating diseases associated with expression of the nucleic acid in humans.
XX Disclosure; Col 47-56; 45pp; English.
XX
XX The invention relates to a compound targeted to specific nucleobases of RecQ protein-like 4 (RECQL4) and which hybridises and inhibits the expression of RECQL4. The compound is useful for inhibiting the expression of RECQL4 in cells or tissues and for treating an animal, particularly a human suspected of having or being prone to a disease or condition associated with expression of RECQL4. The compound is useful for diagnostics, therapeutics and as a research reagent, e.g. prophylactically to prevent or delay infection, inflammation or tumour formation. This sequence represents human RECQL4 polypeptide
XX
XX Sequence 1208 AA;
Query Match 100.0%; Score 1617; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 1.2e-158;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVOALDMPPEAETETLLCYLELHPHHLATYTHCRINCPCGGPAQLOALAHRCPPPL 60
Db 907 QLTVOALDMPPEAETETLLCYLELHPHHLATYTHCRINCPCGGPAQLOALAHRCPPPL 966
QY 61 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRGTGV 120
Db 967 AVCLAQQLPEDPGQSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRGTGV 1026
QY 121 LVFESELAFLHRSFGDLTAEKQICDFLYGRVQARERQALRLRRTFQAFHSAFPCSG 1086

Db 1027 LVFESELAFLHRSFGDLTAEKQICDFLYGRVQARERQALRLRRTFQAFHSAFPCSG 1086
QY 181 PCLEQDEERSTRKDLGRYFEEEGQPGMEDAQGPQARLQDWDQVRCDIRQF 240
Db 1087 PCLEQDEERSTRKDLGRYFEEEGQPGMEDAQGPQARLQDWDQVRCDIRQF 1146
QY 241 LSLRPEKFSRAVARIFHGIGSPCYPAQVYGQDRFRWKYHLHSFHALVGLATEELLQV 300
Db 1147 LSLRPEKFSRAVARIFHGIGSPCYPAQVYGQDRFRWKYHLHSFHALVGLATEELLQV 1206
QY 301 AR 302
Db 1207 AR 1208

RESULT 6
ADQ21562
ID ADQ21562 standard; protein; 1208 AA.
XX AC ADQ21562;
XX 26-AUG-2004 (first entry)
XX Human soft tissue sarcoma-upregulated protein - SEQ ID 4382.
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX Homo sapiens.
XX WO2004048938-A2.
XX 10-JUN-2004.
XX 26-NOV-2003; 2003WO-US038193.
XX 26-NOV-2002; 2002US-0429739P.
XX (PROT-) PROTEIN DESIGN LABS INC.
XX Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.
XX
XX Example 2; SEQ ID NO 4382; 210pp; English.
XX The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytostatic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 1208 AA;
Query Match 100.0%; Score 1617; DB 8; Length 1208;
Best Local Similarity 100.0%; Pred. No. 1.2e-158;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVOALDMPPEAETETLLCYLELHPHHLATYTHCRINCPCGGPAQLOALAHRCPPPL 60

Db 907 QLTVALDMPBEAETLLCYLELPHHWELELLATYTHCLNCPGGAQALQALAHRCPL 966

Qy 61 AVCLAQQLPEDPGGSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRTGV 120

Db 967 AVCLAQQLPEDPGGSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRTGV 1026

Qy 121 LVFSELAFLHRSFGDLTAEEKQICDPLGRVQARERQALRLRTTFOAFHSAVFPSCG 180

Db 1027 LVFSELAFLHRSFGDLTAEEKQICDPLGRVQARERQALRLRTTFOAFHSAVFPSCG 1086

Qy 181 PCLEQQDEERSTRLKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIRQF 240

Db 1087 PCLEQQDEERSTRLKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIRQF 1146

Qy 241 LSLRPEKFSRAVARIPFHGIGSPCYPAQVYGQDRRFRWKYHLHLSFHALVGLATEELLQV 300

Db 1147 LSLRPEKFSRAVARIPFHGIGSPCYPAQVYGQDRRFRWKYHLHLSFHALVGLATEELLQV 1206

Qy 301 AR 302

Db 1207 AR 1208

RESULT 7

ADU80701

ID ADU80701 standard; protein; 1208 AA.

AC ADU80701;

DT 10-FEB-2005 (first entry)

DE Human RTS-associated protein SEQ ID NO 73.

XX anticancer; apoptosis; RecQ; DNA helicase; RNA interference; RNAi;

KW cytostatic; gene silencing; enzyme.

OS Homo sapiens.

XX WO2004100990-A1.

PN 25-NOV-2004.

XX 19-MAY-2004; 2004WO-JP007145.

PF 19-MAY-2003; 2003JP-00140685.

PR (GENE-) GENE CARE RES INST CO LTD.

PA Takagi M, Shimamoto A, Furuichi Y, Sato A;

PI WPI; 2004-833899/82.

XX Apoptosis inducing agent of cancer cell useful as anticancer agent,

DR comprises compound that suppresses expression of RecQ DNA helicase-family

PT gene/function of protein encoded by RecQ DNA helicase-family gene, as

XX active ingredient.

PS Disclosure; SEQ ID NO 73; 262pp; Japanese.

XX This invention describes a novel anticancer apoptosis inducing agent

CC which comprises (a) compound, which suppresses the expression of RecQ DNA

CC helicase-family gene, (b) DNA which expresses the double-stranded RNA

CC with RNA interfering (RNAi) effect with respect to RecQ DNA helicase-

CC family gene, or (c) a compound, which suppresses the function of protein

CC encoded by RecQ DNA helicase-family gene, as an active ingredient. The

CC invention also describes a method for screening for a candidate compound

CC with a protein encoded by RecQ DNA helicase-family gene. A compound,

CC which suppresses the expression of RecQ DNA helicase-family gene is a

CC transduction product or antisense nucleic acid of one part of RecQ DNA

CC helicase-family gene, or a nucleic acid which has ribozyme activity which

CC cleaves the transcription product of RecQ DNA helicase-family gene

CC specifically. It can also be a RecQ DNA helicase-family protein variant

CC which has dominant negative character with respect to the protein encoded

CC by RecQ DNA helicase-family gene, an antibody or a low molecular compound

CC coupled with the protein encoded by RecQ DNA helicase-family gene. The

CC RecQ DNA helicase-family gene is a WRN gene, BLM gene or RecQ1 gene. The

CC method of the invention is useful for manufacturing an anticancer agent

CC as a pharmaceutical composition. The novel agent is highly safe,

CC effective and does not cause any side effects when used as a

CC pharmaceutical. The products of the invention have cytostatic activity.

CC This sequence represents a human helicase-associated RTS protein.

XX

SQ Sequence 1208 AA;

Query Match 100.0%; Score 1617; DB 8; Length 1208;

Best Local Similarity 100.0%; Pred. No. 1.2e-158;

Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QLTVALDMPBEAETLLCYLELPHHWELELLATYTHCLNCPGGAQALQALAHRCPL 60

Db 907 QLTVALDMPBEAETLLCYLELPHHWELELLATYTHCLNCPGGAQALQALAHRCPL 966

Qy 61 AVCLAQQLPEDPGGSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRTGV 120

Db 967 AVCLAQQLPEDPGGSSSVFDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRTGV 1026

Qy 121 LVFSELAFLHRSFGDLTAEEKQICDPLGRVQARERQALRLRTTFOAFHSAVFPSCG 180

Db 1027 LVFSELAFLHRSFGDLTAEEKQICDPLGRVQARERQALRLRTTFOAFHSAVFPSCG 1086

Qy 181 PCLEQQDEERSTRLKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIRQF 240

Db 1087 PCLEQQDEERSTRLKDLLGRYFEEEGQEPGGMEDAQGPFGQARLQDWDQVRCDIRQF 1146

Qy 241 LSLRPEKFSRAVARIPFHGIGSPCYPAQVYGQDRRFRWKYHLHLSFHALVGLATEELLQV 300

Db 1147 LSLRPEKFSRAVARIPFHGIGSPCYPAQVYGQDRRFRWKYHLHLSFHALVGLATEELLQV 1206

Qy 301 AR 302

Db 1207 AR 1208

RESULT 8

ADM90956

ID ADM90956 standard; protein; 1306 AA.

AC ADM90956;

DT 03-JUN-2004 (first entry)

XX Human pharmaceutically useful protein SeqID 349.

DE human; cancer; haematopoiesis; thrombosis; anaemia;

XX cardiovascular disorder; ischaemic heart disease;

KW acute myocardial infarction; respiratory disease; asthma; pneumonia;

KW cystic fibrosis; chronic renal failure; glomerulopathy;

KW gastrointestinal disorder; peptic ulcer; cirrhosis; immune disorder;

KW HIV infection; systemic lupus erythematosus; endocrine system;

KW diabetes mellitus; epilepsy; Alzheimer's disease;

KW amyotrophic lateral sclerosis; skin disorder; cytostatic; anticoagulant;

KW fungal; parasitic; viral infection; vasotropic; antiaesthetic;

KW thrombolytic; antianemic; cardiant; vasotropic; hepatotropic;

KW antiinflammatory; nephrotropic; antiulcer; hepatotropic;

KW immunosuppressive; antiallergic; dermatological; antirheumatic;

KW antiarthritic; antidiabetic; anticonvulsant; neuroprotective; nootropic;

KW antipsoriatic; antibacterial; fungicide; antiparasitic; virucidal;

XX gene therapy; vaccine.

XX OS Homo sapiens.

XX WO2004020595-A2.

PN 11-MAR-2004.

PD 28-AUG-2003; 2003WO-US027107.

XX PF

XX 29-AUG-2002; 2002US-0406576P.
 PR 29-AUG-2002; 2002US-0406611P.
 PR 29-AUG-2002; 2002US-0406612P.
 PR 29-AUG-2002; 2002US-0406616P.
 PR 29-AUG-2002; 2002US-0406640P.
 PR 29-AUG-2002; 2002US-0406655P.
 PR 29-AUG-2002; 2002US-0406666P.
 PR 17-SEP-2002; 2002US-0410946P.
 PR 17-SEP-2002; 2002US-0410951P.
 PR 17-SEP-2002; 2002US-0410953P.
 PR 17-SEP-2002; 2002US-0410957P.
 PR 17-SEP-2002; 2002US-0410960P.
 PR 17-SEP-2002; 2002US-0410962P.
 PR 17-SEP-2002; 2002US-0411019P.
 PR 17-SEP-2002; 2002US-0411022P.
 PR 17-SEP-2002; 2002US-0411024P.
 PR 17-SEP-2002; 2002US-0411037P.
 PR 17-SEP-2002; 2002US-0411046P.
 PR 17-SEP-2002; 2002US-0411052P.
 PR 17-SEP-2002; 2002US-0411082P.
 PR 17-SEP-2002; 2002US-0411111P.
 XX
 PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
 PA (RIKE-) RIKEN INST PHYSICAL & CHEM RES.
 PA (DNAP-) DNAFORM KK.
 XX
 PI Williams LT, Chu K, Lee E, Hestir K;
 XX
 DR WPI; 2004-257410/24.
 DR N-PSDB; ADM90747, ADM91165.
 XX
 PT New human polynucleotides and polypeptides, useful for diagnosing,
 PT preventing and treating proliferative disorders, immune disorders,
 PT cardiovascular disorders, or bacterial, fungal, parasitic and viral
 PT diseases.
 XX
 PS Claim 1; SEQ ID NO 349; 254pp; English.
 XX
 CC This invention relates to novel isolated human polynucleotides and the
 CC encoded proteins thereof. Specifically, it refers to proteases, kinases,
 CC phosphatases, secreted and transmembrane proteins, as well as the derived
 CC peptide fragments, which can be used to develop antibodies and screen for
 CC small molecule agonists and antagonists that can modulate their
 CC activities. The present invention describes polypeptides,
 CC polynucleotides, vectors and host cells useful for diagnosing, preventing
 CC and treating proliferative disorders, e.g. cancer, disorders of
 CC haematopoiesis such as thrombosis and anaemia, cardiovascular disorders,
 CC e.g. ischaemic heart disease and acute myocardial infarction, respiratory
 CC diseases, e.g. asthma, pneumonia or cystic fibrosis, disorders of the
 CC kidney and urinary tract, e.g. chronic renal failure and glomerulopathy,
 CC gastrointestinal disorders, e.g. peptic ulcer or cirrhosis, immune
 CC disorders, e.g. HIV infection and systemic lupus erythematosus, disorders
 CC of the endocrine system, e.g. diabetes mellitus, central nervous system
 CC disorders, e.g. epilepsy, Alzheimer's disease or amyotrophic lateral
 CC sclerosis, skin disorders, e.g. psoriasis, as well as bacterial, fungal,
 CC parasitic and viral diseases. Accordingly, they exhibit many various
 CC activities including cytostatic, anticoagulant, thrombolytic,
 CC antianaemic, cardiant vasotropic, antiasthmatic, antiinflammatory,
 CC nephrotropic, antiulcer, hepatotropic, immunosuppressive, anti-allergic,
 CC dermatological, antiinfective, antiarthritic, antidiabetic,
 CC anticonvulsant, neuroprotective, nootropic, antipsoriatic, antibacterial,
 CC fungicide, antiparasitic and virucidal, such that these polynucleotides
 CC can be used for gene therapy purposes and the development of appropriate
 CC vaccines. This polypeptide is a human protein of the invention.
 XX
 SQ Sequence 1306 AA;
 Query Match 98.6%; Score 1594.5; DB 8; Length 1306;
 Best Local Similarity 92.4%; Pred. No. 2.9e-156;
 Matches 302; Conservative 0; Mismatches 0; Indels 25; Gaps 1;
 QY 1 QLTVALDMPPEE-----AIE TLICYLELHPHHLIATT 35

Db 980 QLTVALDMPPEE GEGEPVSHRGVGLSPRLSPALPPAIE TLICYLELHPHHLIATT 1039
 QY 36 YTHCR LNCPCGPAQLAALAHRCPLAVCLAAQQLPEDPGOGSSSVFDMVKLVDSMGWELA 95
 Db 1040 YTHCR LNCPCGPAQLAALAHRCPLAVCLAAQQLPEDPGOGSSSVFDMVKLVDSMGWELA 1099
 QY 96 SVRRALCOLQWDHEPRTGVRRGTGVLVFESELAFLHRS PGDLTAEEKDQICDFLYGRVQA 155
 Db 1100 SVRRALCOLQWDHEPRTGVRRGTGVLVFESELAFLHRS PGDLTAEEKDQICDFLYGRVQA 1159
 QY 156 RERQALARLRTPQAFHSVAFPSGCGCLEQDDEERSTR LKDLGRYFEEBEGEPGGMED 215
 Db 1160 RERQALARLRTPQAFHSVAFPSGCGCLEQDDEERSTR LKDLGRYFEEBEGEPGGMED 1219
 QY 216 AQGPFGQARLQDWDQVRCDIQFLSLRPEEKFSRAVARIFHGIGSPCYPAQVYGQDR 275
 Db 1220 AQGPFGQARLQDWDQVRCDIQFLSLRPEEKFSRAVARIFHGIGSPCYPAQVYGQDR 1279
 QY 276 RFWRKYLHLSFHALVGLATEELIQQVAR 302
 Db 1280 RFWRKYLHLSFHALVGLATEELIQQVAR 1306
 RESULT 9
 ADM44021
 ID ADW44021 standard; protein; 1216 AA.
 XX
 AC ADW44021;
 XX
 DT 24-MAR-2005 (first entry)
 XX
 DE Mouse RECQL4 protein.
 XX
 KW RECQL4; helicase; protein deactivation; growth disorder;
 KW Rothmund-Thomson syndrome; skin disorder; bone disease;
 KW musculo-skeletal disease; enzyme.
 XX
 OS Mus musculus.
 XX
 PN WO2005001085-A1.
 XX
 PD 06-JAN-2005.
 XX
 PF 25-JUN-2004; 2004WO-JP009380.
 XX
 PR 27-JUN-2003; 2003JP-00185409.
 XX
 PA (AGEN) NAT INST RADIOLOGICAL SCI.
 XX
 PI Abe M;
 DR WPI; 2005-081681/09.
 XX N-PSDB; ADM44020.
 PT Novel non-human mammal, preferably rodent such as mouse exhibiting
 PT characteristics of Rothmund-Thomson syndrome, being RECQL4-gene lacking
 PT mouse having mutation in RECQL4-gene, useful as model of human Rothmund-
 PT Thomson syndrome.
 XX
 PS Disclosure; SEQ ID NO 2; 69pp; Japanese.
 XX
 CC This invention describes a novel knock-out mouse which lacks the RECQL4
 CC gene or has a mutation in one of exons 13 to 22 of the RECQL4 gene
 CC resulting in a loss of helicase activity. The mouse described in the
 CC invention exhibits the characteristics of Rothmund-Thomson syndrome and
 CC is suitable as a model of human Rothmund-Thomson syndrome. The mouse
 CC exhibits growth delay, skin abnormality and bone-formation defects
 CC similar to human Rothmund-Thomson syndrome. This sequence represents
 CC murine RECQL4 which is missing in mice exhibiting Rothmund-Thomson
 CC syndrome.
 XX
 SQ Sequence 1216 AA;

Query Match 70.8%; Score 1144.5; DB 9; Length 1216;
Best Local Similarity 73.2%; Pred. No. 2.1e-109;
Matches 218; Conservative 26; Mismatches 53; Indels 1; Gaps 1;
XX SQ Sequence 1579 AA;
Query Match 26.4%; Score 427; DB 4; Length 1579;
Best Local Similarity 31.1%; Pred. No. 1.8e-34;
Matches 95; Conservative 62; Mismatches 122; Indels 26; Gaps 6;
Qy 1 QLTVALDMPBEATETLLCYLELHPHWHLELLATTYTHCRINCPCGPAQLQALAHRCPPPL 60
Db 916 QSTVALDMPBEATETLLCYLELHPHWHLELLATTYTHCRINCPCGPAQLQALAHRCPPPL 975
Qy 61 AVCLAQQLPEDPGQSSSVFEDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRTGV 120
Db 976 AACQAKWPKDTSQGRSLRFGVWELADSMGKLASVQALHQLKWDPEPKKGAQGTGV 1035
Qy 121 LVSELSAFHLRSPDGLTAEBKQDCDFLYGRVQARERQALRLRTFOAHSHVAFPSCG 180
Db 1036 LVKFSLSAFHLRSGDLTDEEKQDCDFLYRQVAREHKAHLHOMSKAFRSVAFPSCG 1095
Qy 181 PCLEQDDEERSTRKDLALGRYFEREEGQEPGMEDAQPEQCARLQDWEQVRCIRQF 240
Db 1096 PCLEQDDEERSTRKDLALGRYFEREEGQEPGMEDAQPEQCARLQDWEQVRCIRQF 1154
Qy 241 LSLRPEEFSSRAVARIFHGIGSPCYPAQVYQDQRRFRWRKYLHLSFHALVGLATEELL 298
Db 1155 LSLRPEEFSSRAVARIFHGIGSPCYPAQVYQDQRRFRWRKYLHLSFHALVGLATEELL 1212
RESULT 10
ID ABB63317
XX AC ABB63317;
XX DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 16743.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL07420.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 16743; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

Qy 3 TVQALDMPBEATETLLCYLELHPHWHLELLATTYTHCRINCPCGPAQLQALAHRCPPPLAV 62
Db 1290 TVEMLDIPAEINISTLLCTMBLDPFWCISVLSAYVMAKVISYGGPKYLKHAKECPPLAM 1349
Qy 63 CLAQQLPEDP-GQSSSVFEDMVKLVDSMGWELASVRRALCOLQWDHEPRTGVRRTGV 121
Db 1350 AIALQIRDKTFKEDSNIEFSVTDIAAGIGWNSGVVYQLKLEW--VKVNGYFKRSPIT 1407
Qy 122 VESLSAFHLRSPDGLTAEBKQDCDFLYGRVQARERQALRLRTFOAHSHVAFPSCGP 181
Db 1408 VSFYDLGFRIKVPGDFTSEIDNALDLYTRSVKQERTQLIQYVAHGLAAVAYSSCGQ 1467
Qy 182 CLEQQ-DEERSTRKDLALGRYFEREEGQ-----EPGMEDAQPEQCARLQDWEQV 234
Db 1468 CCNADFFQDRGEQKALVRNYFANDYPQDLELETEPSNVPD-----ENII 1512
Qy 235 CDIRQFLSLRPEEFSSRAVARIFHGIGSPCYPAQVYQDQRRFRWRKYLHLSFHALVGLAT 294
Db 1513 DDVHALINMYPDNTFTGTNTARIIFHGIMSPNYPAVINGR-CKFWRAHVKVDFNRIHLAN 1571
Qy 295 EELQ 299
Db 1572 MAIK 1576
RESULT 11
ID ADV13119 standard; protein; 206 AA.
XX AC ADV13119;
XX DT 21-APR-2005 (first entry)
DE Plant full length insert polypeptide seqid 68934.
XX KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX OS Unidentified.
XX PN US2004034888-A1.
XX PD 19-FEB-2004.
XX PF 28-APR-2003; 2003US-00425114.
XX PR 06-MAY-1999; 99US-00304517.
XX PR 05-NOV-2001; 2001US-00985678.
XX (LIUJ/) LIU J.
XX PA (ZHOU/) ZHOU Y.
XX PA (KOVA/) KOVALIC D K.
XX PA (SCRE/) SCREEN S E.
XX PA (TABAV/) TABASKA J E.
XX PA (CAOY/) CAO Y.
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX DR
XX XX

PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.

PS Claim 1; SEQ ID NO 68934; 15pp; English.

XX The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at
 CC ftp://seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
 CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in
 CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomannan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.

XX SQ Sequence 206 AA;

Query Match 9.0%; Score 145.5; DB 8; Length 206;
 Best Local Similarity 25.0%; Pred. No. 2.4e-06;
 Matches 50; Conservative 32; Mismatches 69; Indels 49; Gaps 9;
 QY 123 EFS-ELAFHLRSPG-----DLTAEKDKQICDFLYGRVQARERQALRLRTQAF 171
 DB 29 QFSGEISFELKDPACIYVILKKPDDNALSD-----ITRWLSEVENSKSKLDAMFD-L 82
 QY 172 HSAVFPSC-----GCLIQQDEERSLTKLILGRYFEEEGQEPGGMEDAGQPE 220
 DB 83 ANFAVKGCQRTDGCSGSQTPTCIQK-----IIQVFSKNYSTSDSG--QCTOPQ 129
 QY 221 PGQARLQDQWEDVRCIDIRQFLSLRPEKFSRAVARIFHGICSPCYPAQVYQDQRRWRK 280
 DB 130 RGSFPLQ-----ADIKVFLKNSFAKFTTPRAVARIMHGISSPAFFSATWSKN-HFWGR 181
 QY 281 YLHLSFHVLGLATEELLQV 300
 DB 182 YLEVDFFPVMEAKAELVKL 201

RESULT 12
 ABO80677

ID ABO80677 standard; protein; 762 AA.

AC ABO80677;

XX 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polypeptide #12852.

XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

OS Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nollong J, Deloughery C, Bush D;
 PI WPI; 2003-615309/58.
 XX N-PSDB; ABD14248.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of
 PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 29423; 455pp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABO67836-
 CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html

XX SQ Sequence 762 AA;

Query Match 6.7%; Score 108.5; DB 7; Length 762;
 Best Local Similarity 23.9%; Pred. No. 0.1;
 Matches 72; Conservative 21; Mismatches 79; Indels 129; Gaps 14;
 QY 46 GPAQLQALAH---RCPLAVLCAQLPDPGQSSSVFDMVKLVDSMGWELASVRRALC 102
 DB 470 GPSLGGPRHRPRRRPPLA-----ETPGQRSSAAEH-----APGTGSD----- 507
 QY 103 QLQWDHPTGVRGTGVLVEFSELAFLHLSRPGDLTAEKDKQICDFLYGRVQARERQA-- 160
 DB 508 ---DHGPRDPGR-----DRPGELVQSRPRPSCAAGQGRVHRRLRQADG 547
 QY 161 -----LRLRRTFOAFHSVAFPSGCPCLQEQDEERSTRLKDLILGRYFEEEGQEP 210
 DB 548 GRPRATAGRLPLRLR-----RPGFAPHLPSRTER--RLAQVLGRWRDEPRAP 594
 QY 211 GGMEDA-----QGP-----EPQARLQDQWEDQV 233
 DB 595 GGHRLARRDGHPRRAQRAAMGFLRCVPVHQCGPGDQARGQAALQLPPGLRLDKRAV 654
 QY 234 RCDIRQFLSLRPEEKFSRAVARIFHG-----IGSPCY----PAQVYQDQR 275
 DB 655 QDQLPRRIP-RADRRRRRRAPASAGEGPPATDQPHRRHHPRVGDPHHFGQRPRA---QGRP 710
 QY 276 R 276
 DB 711 R 711

RESULT 13

ABR82253

ID ABR82253 standard; protein; 751 AA.

AC ABR82253;

XX 13-OCT-2003 (first entry)

XX Human nucleic acid-associated protein (NAAP)-Id 1532441CD1.

XX NAAP; nucleic acid-associated protein; cardiant; cytostatic; transgenic;
 KW neuroprotective; gene therapy; human; cancer.

XX

OS Homo sapiens.
XX
XX WO2003052048-A2.
XX
XX 26-JUN-2003.
XX
XX 02-MAY-2002; 2002WO-US014276.
XX
XX 04-MAY-2001; 2001US-0288598P.
XX 17-MAY-2001; 2001US-0291776P.
XX 18-MAY-2001; 2001US-0292172P.
XX 25-MAY-2001; 2001US-0293564P.
XX
XX (INCYTE GENOMICS INC.
XX
XX Yue H, Ding L, Baughn MR, Lal PG, Yue H, Hafalia AJA, Lee EA,
XX Ison CH, Becha SD, Gururajan R, Emerling BM, Griffin JA, Tang YT,
XX Lu DAM, Yao MG, Chawla NK, Ramkumar J, Gandhi AR, Lee SY,
XX Richardson TW, Yang J, Elliott VS, Lu Y, Thangavelu K, He A,
XX Azimzal Y, Raumann BE, Swarnakar A, Burford N;
XX
XX WPI; 2003-541640/51.
XX
XX N-PSDB; ACP355661.
XX
XX New human nucleic acid-associated proteins polypeptide, useful for
XX preparing a composition for diagnosing or treating e.g., cardiovascular
XX or neurological disorders.
XX
XX Claim 1; Page 185-187; 212pp; English.
XX
XX The invention relates to human nucleic acid-associated proteins (NAP)
XX and encoding polynucleotides. The NAP polypeptides can be expressed by
XX standard recombinant methodology. The polypeptides are useful for
XX preparing a composition for diagnosing or treating a disease or condition
XX associated with decreased expression or overexpression of functional NAP
XX e.g., cardiovascular or neurological disorders or cancer. The present
XX sequence represents a human NAP polypeptide
XX
XX Sequence 751 AA;
XX

Query Match 6.6%; Score 107; DB 6; Length 751;
Best Local Similarity 27.0%; Pred. No. 0.14;
Matches 70; Conservative 23; Mismatches 106; Indels 60; Gaps 10;
QY 11 BEAETILCYLHHPHMTLELATYTT-----HCRINCPCGPAQOLAHRCPPPLAVCLA 65
DB 159 EBARERQAAQCPOHGEALRFLCQPCSQLCRECHLD-----PHLDHPCPLAEAVR 210
QY 66 QQLPEDPGQSSSVEFDWVKLVDSMGWEIASVVR-----ALCOLQMD-----HEPRTG 113
DB 211 ARPPGLEG-----LLAGVDNNLVELEAARVKEKALARLQQAARVGTQVEEAAG 261
QY 114 VARGTGVLFESBELAHLRSPGDLTAEBKQICDPLFYGVQ-AREERQALR----- 163
DB 262 VLR--ALTAQKQEVIGQLRAHVAEAAERLAELEGREQVAAAAAFARRVLSIGREA 319
QY 164 -----LRRTPQAFHSVAFSPGCPCLAEQDDEESTRLK---DLIGRYFEERBEQDEG 211
DB 320 EILSLGALIAQRIARLQOGCPWAGPAPCILLPQLEHPRGLDKXNCHILRLSFEBQDPQKXG 379
QY 212 GMEIDA--QGPDEQARLQD 228
DB 380 GKDGACTQGGESQSRRED 398

RESULT 14
ABM80914
ID ABM80914 standard; protein; 755 AA.
XX
XX ABM80914;
XX
XX 18-NOV-2004 (first entry)
XX

DE Tumour-associated antigenic target (TAT) polypeptide PRO81587, SEQ:2360.
XX
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX tumour; diagnosis; cell proliferative disorder; breast cancer;
XX colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX central nervous system cancer; bladder cancer; pancreatic cancer;
XX cervical cancer; melanoma; leukaemia; hybridisation probe;
XX chromosome identification; chromosome mapping; gene mapping;
XX gene therapy; cytostatic.
XX
XX Homo sapiens.
XX
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.
XX
XX 02-OCT-2002; 2002US-0414971P.
XX
XX (GETH) GENENTECH INC.
XX
XX Mu TD, Zhang Z, Zhou Y;
XX
XX WPI; 2004-347921/32.
XX
XX N-PSDB; ACN38689.
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 12; SEQ ID NO 2360; 7273pp; English.
XX
XX The invention relates to human tumour-associated antigenic target (TAT)
XX polypeptides, and their related nucleic acids. The TAT polypeptides are
XX overexpressed in cancer tissues compared to normal tissues, and may thus
XX serve as effective targets for the diagnosis and treatment of cancer in
XX mammals. The invention also relates to nucleic acid and polypeptide
XX sequences at least 80% identical to the TAT nucleic acid and
XX polypeptides; expression vectors and host cells comprising a TAT nucleic
XX acid; an antibody specific for a TAT polypeptide; a peptide or organic
XX molecule which binds to a TAT polypeptide; fusion proteins comprising a
XX TAT polypeptide; and methods and compositions for the treatment or
XX diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
XX antibodies, antagonists, binding molecules and compositions are useful
XX for diagnosing or treating a cell proliferative disorder associated with
XX increased TAT expression, particularly cancers such as breast cancer,
XX colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
XX cancer, pancreatic cancer, cervical cancer, cancers of the central
XX nervous system, melanoma and leukaemia. TAT nucleic acids may further be
XX used as hybridisation probes, in chromosome and gene mapping, in
XX chromosome identification and in gene therapy. The present sequence
XX represents a TAT polypeptide of the invention
XX
XX Sequence 755 AA;
XX

Query Match 6.6%; Score 107; DB 8; Length 755;
Best Local Similarity 27.0%; Pred. No. 0.15;
Matches 70; Conservative 23; Mismatches 106; Indels 60; Gaps 10;
QY 11 BEAETILCYLHHPHMTLELATYTT-----HCRINCPCGPAQOLAHRCPPPLAVCLA 65
DB 159 EBARERQAAQCPOHGEALRFLCQPCSQLCRECHLD-----PHLDHPCPLAEAVR 210
QY 66 QQLPEDPGQSSSVEFDWVKLVDSMGWEIASVVR-----ALCOLQMD-----HEPRTG 113
DB 211 ARPPGLEG-----LLAGVDNNLVELEAARVKEKALARLQQAARVGTQVEEAAG 261
QY 114 VARGTGVLFESBELAHLRSPGDLTAEBKQICDPLFYGVQ-AREERQALR----- 163
DB 262 VLR--ALTAQKQEVIGQLRAHVAEAAERLAELEGREQVAAAAAFARRVLSIGREA 319

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:40:26 ; Search time 11.4 Seconds
(without alignments)
2548.900 Million cell updates/sec

Title: US-09-889-325-4_COPY_907_1208

Perfect score: 1617
Sequence: 1 QLTVALDMPREAIETLLCY.....HLSPHALVGLATELLQVAR 302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	186	11.5	941	2	A86404	probable protein A
2	97.5	6.0	1548	2	T25808	hypothetical prote
3	96.5	6.0	976	2	S40697	processing endopro
4	95	5.9	1582	2	E70876	probable polyketid
5	94.5	5.8	298	2	TC7568	kidney inhibitor o
6	91.5	5.7	1397	2	T46354	hypothetical prote
7	91.5	5.7	2548	2	B59435	myosin IIA [import
8	89	5.5	2274	2	TC0258	adenomatous polypo
9	88.5	5.5	308	2	JC2549	apolipoprotein F p
10	87	5.4	679	2	E71282	probable NH(3)-dep
11	85.5	5.3	548	2	C75499	cytidine/deoxycyt
12	85.5	5.3	634	2	T00359	hypothetical prote
13	85	5.3	637	2	B95878	probable adenylate
14	85	5.3	638	2	JC7753	ring finger B-box
15	84.5	5.2	476	2	T42692	hypothetical prote
16	84.5	5.2	1209	2	T00373	hypothetical prote
17	84.5	5.2	4687	1	A39638	plectin - rat
18	84	5.2	472	2	IS9087	ISG-K54 - human
19	84	5.2	1175	2	IS7549	adenosine deaminas
20	83.5	5.2	343	2	HT0517	probable lppd prote
21	83.5	5.2	574	2	B57072	hypothetical prote
22	83	5.1	547	2	T45635	hypothetical prote
23	83	5.1	948	2	T03225	probable regulator
24	83	5.1	1082	2	T13410	hypothetical prote
25	82.5	5.1	585	2	S49253	rubisco binding pr
26	82.5	5.1	1097	2	G85092	hypothetical prote
27	82.5	5.1	1157	2	H69163	DNA helicase relat
28	82.5	5.1	1545	2	T14288	DNA (cytosine-5-)-
29	82	5.1	1118	2	C95385	probable adenylyate

30	81.5	5.0	1211	2	C83110	probable exonuclea
31	81.5	5.0	2626	2	T31099	myosin-RhoGAP prot
32	81.5	5.0	4574	2	G02520	plectin - human
33	81.5	5.0	4684	2	A59404	plectin [imported]
34	81	5.0	440	1	BVBRCD	cyad protein - Bor
35	81	5.0	473	2	A48949	beta-glucosidase,
36	80.5	5.0	537	2	B87414	RNA polymerase sig
37	80.5	5.0	1013	2	S32214	hypothetical prote
38	80.5	5.0	1689	2	S72467	sodium channel pro
39	80	4.9	507	2	T36009	probable glucose-6
40	80	4.9	1148	2	AD0198	transcription-repa
41	80	4.9	1334	2	B86451	probable copia-cyp
42	79.5	4.9	242	1	RHRUT	thyloliberin precu
43	79.5	4.9	293	1	QOCVPY	B1 protein - pota
44	79.5	4.9	456	2	T39014	hypothetical zinc-
45	79.5	4.9	747	2	I39444	AMP deaminase (EC

ALIGNMENTS

RESULT 1

A86404 probable protein ATP-dependent DNA helicase RecQ [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86404
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucero, J.S.; Maiti, R.; Marzilli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-941 <STO>
A:Cross-references: UNIPARC:UPI000017A6AC; GB:AE005172; NID:g10998928; PIDN:AG26068.1; C:Genetics:

A:Map position: 1

Query Match 11.5%; Score 186; DB 2; Length 941;

Best Local Similarity 24.4%; Pred. No. 1e-07;
Matches 77; Conservative 45; Mismatches 126; Indels 68; Gaps 12;

QY	5	QALDMPREAIETLLCYELHPRHMLLELATYTHGRINPGSPALQALNR---	CPPLA 61
DB	659	QFDDKKEEWMQITLHLEGEVOYLRLPOLNICCTLN-----FHTSTDAQNTIV 707	
QY	62	VELAQQLPEDPQGSSEYFQVWKLVDSWG---WELASVYRALCOLQMDHEBRTGVRGT 118	
DB	708	MLYQSSPTTLAARAIYVAALKSHVKGGLHVPDIPAVASSICVAT-----T 755	
QY	119	GVLYSESELAFLURSPGDLTAEEKD-----QICDFLYGRVQAREFOALRLRR 166	
DB	756	DVLAETQAL-----KGEVYTELKDSAFCYTIKSPKEICSLSHLTWLTETIESCKYRK 809	
QY	167	-TFQAFHSVAFSPGCPGCEQODEESTRLKDLGTYFESEEGQEPQMGDAGPFGQAR 225	
DB	810	LIMMSAAVAALSVSTSLSSGAKQTR--SLQSRFDYFNG-----DEKCDSPSKA- 859	
QY	226	LDQWEDQ---VRCDIQFLSLRPEKSSRAVARIFHGIGSPCYPAQYQDRRFRWKY 281	
DB	860	-----TONCAFPLRADIKVFLQSNRQAKFPPRAIIMHGVSPPAFPNVSWSK-TTFWGRY 913	
QY	282	LHLSPHALVGLATEEL 297	
DB	914	MNVDFVIMEAQTEL 929	

RESULT 2

T25808

hypochemical protein K08B12.5 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 05-Oct-2004

C/Accession: T25808

R/Becker, M.; Mohlmann, P.

submitted to the EMBL Data Library, April 1997

A/Description: The sequence of C. elegans cosmid K08B12.

A/Reference number: Z20091

A/Accession: T25808

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1548 <BEC>

A/Cross-references: UNIPROT:O01583; UNIPARC:UP1000017A3A8; EMBL:U97001; PIDN:AAB52260.1;

C/Genetics:

A/Genetic: CESP:K08B12.5

A/Map position: 5

A/Intons: 39/1; 66/1; 156/3; 278/2; 391/3; 628/1; 1022/2; 1273/3; 1326/1; 1423/2; 1489/

F/956-1005/Domain: protein kinase C zinc-binding repeat homology <K2N>

Query Match 6.0%; Score 97.5; DB 2; Length 1548;

Best Local Similarity 24.8%; Pred. No. 6.3;

Matches 60; Conservative 38; Mismatches 103; Indels 41; Gaps 11;

QY 49 QQLAHRCPLAVCLAQQLPEDPGQSSSVFPMVKLVDSMGWELASVRALCOLQMD 107

Db 499 EIQILNKLDEALAAQOQKPKDEIVASEKKLKEIKERNQVLWKESEIQRELNNIN-D 557

QY 108 HERTGVRCGVLVSEBELAFHLRSPGDLVABED-----QICPLGKRV 153

Db 558 HLDVLVEKAT-VVQQRDDMQAEIADVDLSLTERKSVKRLQDEAKAKQVADF-BEKL 615

QY 154 QAREROALRLR-----TFQAFHS-----VAFSGGPGCEQDERSTRKDL 198

Db 616 KEITEKALIKKQBEVTIERKSVETDDHLSSEVVAAKNTIASIQATNERETIKKLG 675

QY 199 GRVFE---EEGQEPGMEADQGP-EPGQARLQDMEDVRCDIHQFLSLRPE-EKFSRA 253

Db 676 QRMDEERASHQAQSEQEKQLEAHYERAKQKMLQDMVEGMVNER---GLRDEIKLSQGM 732

QY 254 VA 255

Db 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

QY 733 AA 734

RESULT 4

E70876

probable polyketide synthase pks4 - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004

C/Accession: E70876

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulterson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98255987; PMID:9634230

A/Accession: E70876

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-1582 <COL>

A/Cross-references: UNIPROT:O50437; UNIPARC:UP1000003ADC; GB:AL010186; GB:AL123456; NID:

A/Experimental source: strain H37RV

C/Genetics:

A/Genetic: pks4

C/Superfamily: Mycobacterium tuberculosis probable polyketide synthase pks4; acyl carrier

ology

C/Keywords: carrier protein

F/35-315/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

F/1242-1422/Domain: short-chain alcohol dehydrogenase homology <SADH>

F/1506-1578/Domain: acyl carrier protein homology <ACP1>

Query Match 5.9%; Score 95; DB 2; Length 1582;

Best Local Similarity 22.2%; Pred. No. 10;

Matches 58; Conservative 29; Mismatches 72; Indels 102; Gaps 13;

QY 44 PGCPAQQLAHRCPLAVCLAQQLPED-----PGQSSSVFDM 83

Db 791 PGEPRLFPVYTRSA-----ASVLPSDLANLQAGLRGMRYDSHPHGLKATLVDN 843

QY 84 VKLVDSMGWELASVRALCOLQ-WDHEPRTVRG-----TGLVFEFS 125

Db 844 DEVTVAAL---VAS-----QLQSGSQDEDEAVMNGIWTARLRGRLRPAERRTAVVEYR 894

QY 126 E--LAFHLRSPGDLTAEEKQICDPLXGVQAREQALRLRTFOAFHSVAPPSGCPCL 183

Db 895 RDGRLQIRTPGDLSEIE-----FVTFDRVA-FGPGE-I 926

QY 184 EQQDERSTRKDL---GRY-----FEERGGQEPGMEADQGPPEGQARLQD----- 228

Db 927 EVAVTASVAVFADVLVAFGRYPTFEGTRQQLGIDFAGVTVAVGVDYTHRIIDHVGMSA 986

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

QY 229 ---WEDQVRCDIRQFLSLRPE 246

Db 987 NGCMSTFVRCDARLAVTLRPE 1007

RESULT 5

UC7568
Kidney inhibitor of apoptosis protein - human
C/Species: Homo sapiens (man)
C/Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C/Accession: UC7568
R/Lin, J.H.; Deng, G.; Huang, Q.; Morser, J.
Biochem. Biophys. Res. Commun. 279, 820-831, 2000
A/Title: KIAA, a novel member of the inhibitor of apoptosis protein family.
A/Reference number: UC7568; MUID: 21092523; PMID:11162435
A/Contents: Fetal kidney
A/Accession: UC7568
A/Molecule type: mRNA
A/Residues: 1-298 <LIN>
A/Cross-references: UNIPROT:Q96CA5; UNIPARC:UPI00001269A6
C/Comment: This protein, a new member of the inhibitor of apoptosis protein family, plays a role in the regulation of apoptosis.
C/Genetics:
A/Gene: kiap
A/Map position: 20q13.3
C/Keywords: apoptosis

Query Match 5.8%; Score 94.5; DB 2; Length 298;
Best Local Similarity 20.8%; Pred. No. 1.6;
Matches 62; Conservative 32; Mismatches 113; Indels 91; Gaps 12;

QY 25 PNNMELATTYTHGRINCPGPPAQLQALAHRCPLAVCLAQQLPEPDGSSSVFPMV 84
DB 16 PSHMA-----AGDPTQ-----ERCGRPSL-----GSPVGLDTC 45
QY 85 KLVDSGWEVLASVRALCOLQMDHEPRTGVRGTG-----LVFSELAFLHSPG 135
DB 46 RMDHVDGIIQLQRLTTEEBEGBGATLSRGPAPGSGSEBRLASVYDMPPLTAEPV 105
QY 136 DITAE-----EKQOI-CDPLYGRVQARER-----QALALR-----RT 167
DB 106 ELLAAAGPFTHTGQDKVRCFCYCGIQSWKRGDPWTEHAKVFPSCQFLIRSGKGRDFVHS 165
QY 168 FQAFHVAAPSCGCLCEQODDERSTRKLLG-----RYEEBEGGPGMEDAQP 219
DB 166 VQETHSQLIGSDWPMBEPDAAPVAPSVASGYPPLPTPRREVQSSAQPGVSPAEAO 225
QY 220 -----EPGOARLQMDQVRCDIROFLSLRPEKFSRAVARIF-----HGIGSPCY 267
DB 226 RAMWVLEPFGAR-----DVEAQLRLQSEKTKVCLDAVSLVFPCHLYCAECAP 277

RESULT 6

T46354
hypothetical protein DKFZp434F1016.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C/Accession: T46354
R/Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A/Reference number: Z23037
A/Accession: T46354
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1397 <AAA>
A/Cross-references: UNIPROT:Q9NTG2; UNIPARC:UPI000007143E; EMBL:AL137287
A/Experimental source: adult testis; clone DKFZp434F1016
C/Genetics:
A/Note: DKFZp434F1016.1
F/849-897/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 5.7%; Score 91.5; DB 2; Length 1397;
Best Local Similarity 21.5%; Pred. No. 18;
Matches 55; Conservative 38; Mismatches 76; Indels 87; Gaps 12;
QY 47 PAQLQALAHRCPLAVCLAQQLPEDPG-----QGSSVFPDMVKLVDSM 90
DB 647 PDELAAY-HPTPLS-----PELPGSCRKEFKENKPSPAKRRKRSVKSVAL-DSM 697

QY 91 GWEIASVVRALCOLQMDHEPRTGVRGTGVLVEFSEL-----AFHLRSPGDLTAEE--KDQ 144
DB 698 HWQNDSVQ-----IIASVDSLKSMDEFLTKKYNLDLNDSDSKDT 736
QY 145 ICDPLYGRVQARERQALRLRTQAFHSAVFPSCGCLCEQODDERSTRKLLGRYEE 204
DB 737 LVDPVF-----KKALKEFRQNIFFSYSSAL-----AMDGKSTRYKDLVAFEQI 781
QY 205 EEGEPGGMEDAQGPBPQARLQMDQVRCDIROFLSLRPEKFSRAVARIFHGIGSP 264
DB 782 LEKTRMLEQRDSLSESP-----VRVAVNTFKVFLDEYAN-----EFKT-----SD 821
QY 265 CYPAGVYGQDRRFWRK 280
DB 822 CTATKVPKTERKKRRK 837

RESULT 7

E59435
myosin IXA [imported] - human
C/Species: Homo sapiens (man)
C/Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 09-Jul-2004
C/Accession: E59435
R/Gorman, S.W.; Halder, N.B.; Griehammer, U.; Swiderski, R.E.; Kim, E.; Welch, J.W.; See
Genomics 59, 150-160, 1999
A/Title: The cloning and developmental expression of unconventional myosin IXA (MYO9A) a
A/Reference number: E59435
A/Accession: E59435
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2548 <GOR>
A/Cross-references: UNIPROT:Q9UNJ2; UNIPARC:UPI00002706F; GB:NP_008832; PTD:95902012; P1

Query Match 5.7%; Score 91.5; DB 2; Length 2548;
Best Local Similarity 21.5%; Pred. No. 37;
Matches 55; Conservative 38; Mismatches 76; Indels 87; Gaps 12;

QY 47 PAQLQALAHRCPLAVCLAQQLPEDPG-----QGSSVFPDMVKLVDSM 90
DB 1798 PDELAAY-HPTPLS-----PELPGSCRKEFKENKPSPAKRRKRSVKSVAL-DSM 1848
QY 91 GWEIASVVRALCOLQMDHEPRTGVRGTGVLVEFSEL-----AFHLRSPGDLTAEE--KDQ 144
DB 1849 HWQNDSVQ-----IIASVDSLKSMDEFLTKKYNLDLNDSDSKDT 1887
QY 145 ICDPLYGRVQARERQALRLRTQAFHSAVFPSCGCLCEQODDERSTRKLLGRYEE 204
DB 1888 LVDPVF-----KKALKEFRQNIFFSYSSAL-----AMDGKSTRYKDLVAFEQI 1932
QY 205 EEGEPGGMEDAQGPBPQARLQMDQVRCDIROFLSLRPEKFSRAVARIFHGIGSP 264
DB 1933 LEKTRMLEQRDSLSESP-----VRVAVNTFKVFLDEYAN-----EFKT-----SD 1972
QY 265 CYPAGVYGQDRRFWRK 280
DB 1973 CTATKVPKTERKKRRK 1988

RESULT 8

T30258
adenomatous polyposis coli protein 2 - mouse
N/Alternate names: APC2 protein
C/Species: Mus musculus (house mouse)
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C/Accession: T30258
R/van Be, J.H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Miles, A.; Kuipers, J.
Curr. Biol. 9, 105-108, 1999
A/Title: Identification of APC2, a homologue of the adenomatous polyposis coli tumour sui
A/Reference number: Z20796; MUID:99147086; PMID:10021369
A/Accession: T30258
A/Status: preliminary
A/Molecule type: DNA

A:Residues: 1-2274 <VAN>
A:Cross-references: UNIPROT:Q9Z1K7; UNIPARC:UPI00000296CC; EMBL:AJ130783; NID:g4210431;
A:Gene: APC2
A:Introns: 47/3; 78/1; 138/2; 174/3; 212/3; 238/3; 271/3; 396/1; 428/1; 474/3; 500/3; 53

Query Match 5.5%; Score 89; DB 2; Length 2274;
Best Local Similarity 21.9%; Pred. No. 52;
Matches 68; Conservative 42; Mismatches 113; Indels 88; Gaps 16;

QY 48 AQLALAHRCPLAVCLAQQLPEDEP---GQSSSVFED---WKLVDMSMGW---LA 95
DB 84 SSLYNLKFAHPALPEPAPAPPCGSPVHGSPKDSFGLSRATIRLLELDQERCFLLS 143

QY 96 SVRALQQLQMDHPRGVGRGTGL-----VERSELAFLRSPGDILTA 139
DB 144 EIEEKEKMYYSQLOQLSLRDELPHVDTFSMQMDLIRQOLRE---AQHRS---LM 197

QY 140 EKKDQICDFLYGRVQARERQALRLRTFOAFHSAVFPSCPCLEQODEERSTRLKDLG 199
DB 198 EERGTDDEMYQKQIR---ASRLQIDKEL-----LEAQDRVQOTEPQALLA 242

QY 200 ---RYEEBEGQEPGEMDAQEPGQARLQ-----DWEQVCDIRQFLSL 243
DB 243 VKPVAVEEGQEAIVPTHEDEGT-PQPGNSKVEVFWLLSMLATDQEDTATTLA--MSS 299

QY 244 RPEKPSR-----AVARITHG-----IGSPCTPAQVYGDDRPFWRKYLHSFHALV-- 290
DB 300 SPESCVMMRRSGCLPLLQILHTEAGSVGRAGIPGARGADARM---RANVALHNTVFS 356

QY 291 ----GLATEEL 297
DB 357 QPDGLARKEM 367

RESULT 9
JC2549
apolipoprotein F precursor - human
C:Species: Homo sapiens (man)
C>Date: 26-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 09-Jul-2004
C:Accession: JC2549; PC2384
R:Day, J.R.; Alberts, J.J.; Gilbert, T.L.; Whitmore, T.R.; McConathy, W.J.; Wolfbauer, G.
Biochem. Biophys. Res. Commun. 203, 1146-1151, 1994
A:Title: Purification and molecular cloning of human apolipoprotein F.
A:Reference number: JC2549; MUID:94380022; PMID:8093033
A:Accession: JC2549
A:Molecule type: mRNA
A:Residues: 1-308 <DAY>
A:Cross-references: UNIPROT:Q13790; UNIPARC:UPI0000125C59; GB:L27050
A:Accession: PC2384
A:Molecule type: protein
A:Residues: 147-176;233-247 <DA2>
A:Cross-references: UNIPARC:UPI0000179D12; UNIPARC:UPI0000179D13
C:Genetics:
A:Gene: GDB:APOF
A:Cross-references: GDB:391034; OMIM:107760
A:Map position: 12pter-12qter
C:Keywords: glycoprotein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:127-146/Domain: propeptide #status predicted <PRO>
F:147-303/Product: apolipoprotein F #status predicted <MAT>
F:164-184/Domain: hydrophobic #status predicted <HY2>
F:207-227/Domain: hydrophobic #status predicted <HY3>
F:274-294/Domain: hydrophobic #status predicted <HY4>
F:100,121,249/Binding site: carbonyl (Asn) (covalent) #status predicted

Query Match 5.5%; Score 88.5; DB 2; Length 308;
Best Local Similarity 23.7%; Pred. No. 5.3;
Matches 71; Conservative 34; Mismatches 103; Indels 91; Gaps 16;

QY 14 IETLLCYLEIHP-----HHWELIATVYTHCLNCPG--GPAQLQALAHRC 58
DB 4 VELLCTVLLHPVATSVGKQTVNLMPPLSLSESGTSSDPLSCQFLHPSLPSFSMAP 63

QY 59 ----PLAVCLAQQLPEPDGSSSVFEDMYKLVDMSMGWELASVRALCOLQMDHEPRTGV 114
DB 64 LPEKVLSTALRNALAE-----AGCOADV-----WAL-----QLQL---YRQGG 98

QY 115 RRGVAVVERSELAFLRSPGDLTAEKQDQICDFLYGRVQ--AREQALRLRRTF----- 168
DB 99 VNAIVQVLIQ-----HNRGIQKGRSTERNVSEVALQALQALAREQOSTGRVGRSLPTED 152

QY 169 -----QAFHSAV--FPSCGP-----CLEQODEERSTRLKDLGRYPFEE 206
DB 153 CENEKQAVANVQVLLRGVGFYVNLGTALYATQNCIGKARERRDQADILDGYDLMTMA 212

QY 207 GQE--PGEMDAQEPG-----QARLQMDQVRCIRQFLSLRPE-EKFSAPAVARI 257
DB 213 GMSGGPMGLAISALKEALRSVQQLIYVQDDKDAVISQ-----PETTEGGRALISDV 266

RESULT 10
E71282
probable NH(3)-dependent NAD(+) synthetase (nade) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004
C:Accession: E71282
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
tson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDo
rthy, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: E71282
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-679 <COL>
A:Cross-references: UNIPROT:O83759; UNIPARC:UPI000012PC89; GB:AE001249; GB:AE000520; NID:
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0780

Query Match 5.4%; Score 87; DB 2; Length 679;
Best Local Similarity 24.6%; Pred. No. 18;
Matches 44; Conservative 10; Mismatches 67; Indels 58; Gaps 7;

QY 31 LIAATTYHC-----RLNCPGGRPAQLQALAHRCPL--AVCLAQQLPEPDGQ----- 75
DB 297 IFASAKAHATLRBRPRVPCPPPAFOKSDAVPPLTGAVCLAVASAPSDTDGFLQRTIDL 356

QY 76 -----SSVFERDWMKLVDSMGWELASVRALCOLQMDHEPRTGV 114
DB 357 AAQGVALLRLEHMGCRLLVGVSGVDSACALLICARALDPLSTART--QLYALTLPGFGT 414

QY 115 RRGVAVVERSELAFLRSPGDLTAEKQDQICDFLYGRVQARER 158
DB 415 TSGTKGAAGQAFARALGCTVQGISASAVTHNLHDIGHITM-----QQCDGYENNAQARER 468

RESULT 11
C75499
cytidine/deoxycytidylate deaminase/nudix/methyltransferase domains protein - Deinococcus
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: C75499
R:White, O.; Bisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; f
M.; Shen, M.; Vamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75499
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-548 <WHI>
A:Cross-references: UNIPROT:Q9BRM3; UNIPARC:UPI000003D09; GB:AE001918; GB:AE000513; NID:

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0603

A:Map position: 1

Query Match 5.3%; Score 85.5; DB 2; Length 548;

Best Local Similarity 22.1%; Pred. No. 19;

Matches 73; Conservative 41; Mismatches 135; Indels 81; Gaps 16;

8 DMEPAITLTCYELHPHMLE-LIATTTTCRLNCGGPAQ--LQALHRCP-PLAVC 63

61 DLAAHAIALLSVPLRRPESLMTVLTVEPCP-QCGAVAVMSGIRGSAAPPWGQC 119

64 LAQOLPEBPGGSSVPEPDMVLTVDMSGWELEA---SVRALCOLMDHEPTGVRRGTG 119

120 -ARLITDDP-----YSSKGRVSRAPBEPLOKALRLMLVALLEGRPREBR 165

120 VLVERSELAFLHRSQDL-----TAEKQDQICPLGRVQARERQALRLRT 167

166 ILQFSRYKADIKARELHAGTLARLRSGAGLEBALTELLGALPLEMLVLELS-- 223

168 FOAFHVAFPSCGPCIQODERSTRKDLGR---YFEEERG-----QEPG 212

224 -PARRTAFAPDLSPGLER-----TGRACAMIEREDGFVLTARGTGTLPGGG 270

213 MEDAGPREGQARLDWED-QYRCIDRQFLSLRPEKFSRAVARIFHGIGSPCYPAQY 271

271 IEPGETPEGAAYR-EMEEVGARCEV-----AGAGWTLDDGSSVCVPLRVL 316

272 G-QDRRFMRKYHLSPHALVGLATEELLQV 300

317 TLESSEGRPLIWNPHALPMADVDQLROY 346

RESULT 12

T00359

hypothetical protein KIAA0680 - human

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004

C:Accession: T00359

R:ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.

DNA Res. 5, 169-176, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete

A:Reference number: Z14142; MUID:98403880; PMID:9734811

A:Accession: T00359

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-634 <ISH>

A:Cross-references: UNIPROT:O75167; UNIPARC:UPI000006E0AA; EMBL:AB014580; NID:93327173;

A:Experimental source: brain; clone HK02746

C:Genetics:

A:Note: KIAA0680

Query Match 5.3%; Score 85.5; DB 2; Length 634;

Best Local Similarity 22.7%; Pred. No. 23;

Matches 51; Conservative 26; Mismatches 91; Indels 57; Gaps 9;

58 PPLAVCLAQOLPEBPGGSSVPEPDMVLTVDMSGWELEAVRRALCOLQMDHEP-RTGYR 115

327 PFAVAPASPLAPPLPLEDQCITASDPVVLVSAGADLPVGLDPSQLMAEPTNRTTLY 386

116 RGVGVV-----FSELAFHLSRQDLTAERKQDQICPLGRVQARERQALR 163

387 SGTGLSVNENNAKCFITKEELGRTVPQLTPG-LMGSSSSP-----SASEDG--- 434

164 LRRTFOAFHVAFPSCGPCIQODERSTRKDLGRYFEEERGQEPGME----- 215

435 -HREYQANS---DSGPILYTDDDED-----EDDGSGSALASKIRRDTL 479

216 --AAGPEPGARLD-----WEDQVRCIDRQFLSLRPEKFSR 252

480 AITLGNRPSPKCELDKNIILQRTSEERQELRQOIGTGLVRLSOR 524

RESULT 13

B95878

probable adenylate cyclase (EC 4.6.1.1) [imported] - Sinorhizobium meliloti (strain 1021)

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004

C:Accession: B95878

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmeier, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: B95878

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-637 <KUS>

A:Cross-references: UNIPROT:Q92WN9; UNIPARC:UPI00000CB49C; GB:AL591985; PIDN:CAC48690.1;

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; LeJaure,

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: cyar7; SMD20300

A:Genome: plasmid

C:Keywords: phosphorus-oxygen lyase

Query Match 5.3%; Score 85; DB 2; Length 637;

Best Local Similarity 21.9%; Pred. No. 25;

Matches 52; Conservative 30; Mismatches 77; Indels 78; Gaps 10;

51 QALAHRCPLAVCLAQOLPEBPGGSSVPEPDMVLTVDMSGWELEAVRRALCOLQMD-HE 109

397 RALLHRA-----LELDGYAAARANLCMTIYVFPV-QNLSGRATTTDVGTSSE 444

110 PRTYGRGTGVVERSELAFLHRSQDLTAERKQDQICPLGRVQARER----- 158

445 ABOAVRLDPLNLAAQGVQVLSFGLSATGDYP-----GAMQAAQRAVELNPDPDS 492

159 -QALARLRTFOAFHVAFPSCGPCIQODERSTRKDLGRYFEEERGQ---EPGME 214

493 LMAALAAQVRFESYDAV-----QNAERARLRHPMAEYITTYVGGALYAAGRUD 542

215 DAQGEPPGARLDWEDQV--RCDIR-----QFLSLRPEKFSRAVARI 257

543 EA-----DEVLRRCILRAPQEAQCLLRVAVLSQRGVDEGAQRTRMARL 585

RESULT 14

UC7753

ring finger B-box coiled-coil protein, GOA - human

C:Species: Homo sapiens (man)

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004

C:Accession: UC7753

R:Vandepuute, D.A.A.; Melje, C.B.; van Darstel, M.; Ileenstra, S.; Tolst-Keizers, H.; Das,

Biochem. Biophys. Res. Commun. 286, 574-579, 2001

A:Title: GOA, a novel gene encoding a ring finger B-box coiled-coil protein, is overexpres

A:Reference number: UC7753; MUID:21402356; PMID:11511098

A:Accession: UC7753

A:Molecule type: DNA

A:Residues: 1-638 <VAN>

A:Cross-references: UNIPROT:O96LD4; UNIPARC:UPI000013CE62; GB:AY026763

C:Comment: This protein, a member of the subfamily of the ring finger proteins, has an in

as well.

C:Genetics:

A:Gene: goa

A:Map position: 17q24-25

Query Match 5.3%; Score 85; DB 2; Length 638;

Best Local Similarity 22.6%; Pred. No. 25;
Matches 74; Conservative 29; Mismatches 104; Indels 120; Gaps 18;

QY 47 PAQIQALAHRCPPPLAVCLIAQQLPEDPGQSSSVFEDMVKL-----VDSM 90
Db 183 PRHRLPERYCRARVCLCEACAAQEHGHELVPLEGERALQEAQSKVLSAVEDRMDEL 242
QY 91 GWELASVRRALCOLQMDH-EPRITGVRR-----GTGYL--VFSELAFLHRLSPG 135
Db 243 GAGIAQSRRTVALIKSAVAERERYRLPADAAAALQGFOTVLGFTIEGEMAMIGRSQG 302
QY 136 DLTAEEKDQICDPLYGRVQAREQALRLRRTFQAF--HSVAF-----PSCG 180
Db 303 DL-----RRQEQRSLSRARQNLGVPKADSVSLQELALRLALMEDCG 348
QY 181 PCLREQDDEERSTR-----LKDLIG---RYFEEBEGQEPGEMDAQPE---PGQAR 225
Db 349 PGPGPPRELSPFKSSQAVRAVRDMLAVACVQWELRG--PGNED--GPQKLDSEADAE 404
QY 226 LQDWE-----DQYRCDIROFLSLRPEKFSRAVARIFHGIG 262
Db 405 PQLDESTNLESEARPDYFLKPAIYVDLSDPTAD--KFLQL-----FGTKGVKRVLCPIN 457
QY 263 SPCTPA-----QYVGQ---DR--RFW 278
Db 458 YPLSPTRPTHCEQVLGEGALDRGTYW 484

RESULT 15

T42692
hypothetical protein DKFZp434D0428.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C/Accession: T42692
R/Bioecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A/Reference number: Z22230
A/Accession: T42692
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-476 <AAA>
A/Cross-references: UNIPROT:Q9UFA3; UNIPARC:UPI0000070625; EMBL:AL133088
A/Experimental source: adult testis; clone DKFz434D0428
C/Genetics:
A/Note: DKFz434D0428.1

Query Match 5.2%; Score 84.5; DB 2; Length 476;
Best Local Similarity 22.2%; Pred. No. 19;
Matches 69; Conservative 26; Mismatches 117; Indels 99; Gaps 13;

QY 3 TVQALDMEBEAIEITLCTLEBH-----PHW--LELLATYTHCRLNCPGPAQL 50
Db 5 TVDTLLASKSSKSCAGILCRTLHLBELQPLQRPSPWPQASILGATVTVLRL-CDGSAAPA 63
QY 51 QAL-AHRCPPPLAVCLIAQQLPEDPGQSSSVFEDMVKLVDSMGWELASVRRALCOLQMDHE 109
Db 64 SSVGGHLCGTIAGCVRVQ-----RAALDF-----LGTLSQGTG 96
QY 110 PRIGVRGTVGVLFSELAFLHRLSPGDLTAEEKDQICDPLYGRVQAREQALRLRRTFQ 169
Db 97 PQELVTQALAVLLELC-----LESPPGSSPTVLKKAFOATLAWMLSSPKTPGCSDLGPLIP 150
QY 170 AFHSVAF-----SCGFLREQDDEERSTRLDKLGRYFEEBEGQEPGEMD-----AQ 217
Db 151 QFLNELFPVLQKRLCHPCWEVRDS--ALEFLTQLSRHW-----GGQADFRCALTASE 200
QY 218 GPERGQARLQDWMEDQVR-----CDIROFLSLR-----PEE 247
Db 201 VFQALALQLQDPEBSYVRASAVTAMGQLSSQGLHAPTSEHAERQSLFLELLHITLSVDSB 260
QY 248 KFSRAVARIRF 258
Db 261 GPPRAVMQVF 271

Search completed: December 27, 2005, 21:54:08
Job time : 14.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:39:51 / Search time 58.8 Seconds
(without alignments)
3623.632 Million cell updates/sec

Title: US-09-889-325-4_COPY_907_1208

Perfect score: 1617
Sequence: 1 QLTQALMPEALFTLLCY.....HLSPHALVGLATRELLQVAR 302

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1617	100.0	1208	1 RECO4_HUMAN	094761 homo sapien
2	1144.5	70.8	1216	1 RECO4_MOUSE	07507 mus musculu
3	833	51.5	1500	2 Q4JX8_XENLA	04jnx8 xenopus lae
4	753.5	46.6	910	2 Q4RLC3_TETNG	04rlc3 tetraodon n
5	464.5	28.7	1058	2 Q7PM8_ANOCA	07pm8 anopheles g
6	427	26.4	1530	2 Q9NH11_DROME	09nh11 drosophila
7	427	26.4	1579	2 Q9VSE6_DROME	09vse6 drosophila
8	186	11.5	941	2 Q9C6N0_ARATH	09c6n0 arabidopsis
9	184.5	11.4	870	2 Q8W028_ARATH	08w028 arabidopsis
10	169	10.5	874	2 Q7FPH0_ORYSA	07fph0 oryza sativ
11	169	10.5	927	2 Q7X829_ORYSA	07x829 oryza sativ
12	107	6.6	755	1 TR156_HUMAN	09bxr2 homo sapien
13	107	6.6	809	2 Q6ED36_HUMAN	09ed36 homo sapien
14	107	6.6	900	2 Q6NR70_HUMAN	06nr70 homo sapien
15	107	6.6	1029	2 Q6UX74_HUMAN	06ux74 homo sapien
16	107	6.6	1377	2 Q9P2A8_HUMAN	09p2a8 homo sapien
17	106.5	6.6	1166	2 Q96JUA_HUMAN	096j14 homo sapien
18	106.5	6.6	1681	2 Q8ND48_HUMAN	08nd48 homo sapien
19	106	6.6	253	2 Q6ZWH8_HUMAN	06zwh8 homo sapien
20	105	6.5	966	2 Q72357_THET2	072357 thermus the
21	102.5	6.3	447	2 Q753C3_ASHEO	0753c3 ashyia goss
22	101.5	6.3	4678	2 Q46SH2_TETNG	046sh2 tetraodon n
23	101	6.2	966	2 Q5S1S6_THET8	05s1s6 thermus the
24	98	6.1	678	1 K0323_HUMAN	015037 homo sapien
25	97.5	6.0	1592	2 C01583_CAEBL	001583 caenorabdi
26	97	6.0	914	2 Q6WR18_DROTK	06wr18 drosophila
27	96.5	6.0	976	1 XPR6_YARLI	042781 yarrowia li
28	96.5	6.0	1418	2 Q5R129_BRABE	05r129 brachydanio
29	96	5.9	484	2 Q81Z03_HUMAN	081z03 homo sapien
30	95.5	5.9	563	2 Q5KG20_CRYNE	05kg20 cryptococcu
31	95.5	5.9	655	2 Q9FBR7_STRCO	09fbr7 streptomyce

32	95.5	5.9	1469	2 Q6PD04_MOUSE	06pd04 mus musculu
33	95.5	5.9	1517	2 Q6KAS1_MOUSE	06kas1 mus musculu
34	95	5.9	1582	2 O50437_MYCTU	050437 mycobacteri
35	95	5.9	2085	2 Q7U0G2_MYCBO	07u0g2 mycobacteri
36	95	5.9	2101	2 Q8VKS2_MYCTU	08vks2 mycobacteri
37	94.5	5.8	298	1 B1RC7_HUMAN	096ca5 homo sapien
38	94.5	5.8	498	2 Q7X3G4_THIRO	07x3g4 thicapsa r
39	94.5	5.8	619	2 Q4K3N4_PSEPF5	04k3n4 pseudomonas
40	94	5.8	570	2 Q55RL5_CRYNE	055rl5 cryptococcu
41	94	5.8	638	2 Q7U8M5_SYNXP	07u8m5 synchococc
42	94	5.8	798	2 Q5YTF9_NOCFA	05ytf9 nocardia fa
43	93.5	5.8	356	2 Q7S474_NEUCR	07s474 neurospora
44	93.5	5.8	610	2 Q41XN8_AZOVI	041xn8 azotobacter
45	93.5	5.8	1557	2 Q4S161_TETNG	04s161 tetraodon n

ALIGNMENTS

RESULT 1	RECO4_HUMAN	STANDARD;	PRT; 1208 AA.
AC	094761; Q96DW2; Q96F55;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	13-SEP-2005 (Rel. 48, Last annotation update)		
DE	ATP-dependent DNA helicase Q4 (EC 3.6.1.-) (RecQ protein-like 4)		
DE	(RecQ4) (RTS).		
GN	Name=RecQ4; Synonym=RECQ4;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;		
OC	Homo		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE [mRNA], AND TISSUE SPECIFICITY.		
RC	TISSUE=Testis;		
RX	MEDLINE=99097344; PubMed=9878247; DOI=10.1006/geno.1998.5595;		
RA	Kitao S., Ohnogi I., Ichikawa K., Goto M., Furuchi Y., Shimamoto A.;		
RT	"Cloning of two new human helicase genes of the RecQ family:		
RT	biological significance of multiple species in higher eukaryotes.;"		
RL	Genomics 54:443-452(1998).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE [GENOMIC DNA], SUBCELLULAR LOCATION, AND DISEASE.		
RC	TISSUE=Lymph, and Placenta;		
RX	MEDLINE=2288257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,		
RT	Klausner R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RT	Altehouli S.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Browstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bosa S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Murny D.W., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whitling M., Madan A., Young A.C., Green E.D., Dickson M.C.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,		
RA	Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,		
RA	Scharch A., Schein J.E., Jones S.J.M., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences.;"		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		

[4]
 RN INDUCTION
 RP PubMed=11032027; DOI=10.1038/sj.onc.1203841;
 RX Kanabe T., Tsuyama N., Kiteo S., Nishikawa K., Shimamoto A.,
 RA Shiraori T., Matsunoto T., Anno K., Sato T., Mitsuhi Y., Seki M.,
 RA Shamoto T., Goto M., Ellis N.A., Ide T., Furiuchi Y., Sugimoto M.,
 RT "Differential regulation of human RecQ family helicases in cell
 RT transformation and cell cycle.";
 RL Oncogene 19:4764-4772(2000).
 RN DISEASE.
 RP PubMed=12952869; DOI=10.1093/hmg/ddg306;
 RA Siitonen H.A., Kopra O., Kaeerlaeinen H., Haravuori H., Winter R.M.,
 RA Saeenenen A.-M., Peltonen L., Kesilae M.,
 RT "Molecular defect of RAPADILINO syndrome expands the phenotype
 RT spectrum of RECQL diseases.";
 RL Hum. Mol. Genet. 12:2837-2844(2003).
 RN INTERACTIONS WITH UBR1 AND UBR2, SUBCELLULAR LOCATION, FUNCTION, AND
 RP IDENTIFICATION BY MASS SPECTROMETRY.
 RX PubMed=15317757; DOI=10.1093/hmg/ddh269;
 RA Yin U., Kwon Y.T., Varshavsky A., Wang W.,
 RT "RECQL4, mutated in the Rothmund-Thomson and RAPADILINO syndromes,
 RT interacts with ubiquitin ligases UBR1 and UBR2 of the N-end rule
 RT pathway.";
 RL Hum. Mol. Genet. 13:2421-2430(2004).
 RN [7]
 RP VARIANTS ASP-267 AND GLN-1005.
 RX PubMed=12601557; DOI=10.1007/s100380300016;
 RA Roverai G., Beghini A., Zamburo G., Paradisi M., Larizza L.,
 RT "Identification of two novel RECQL4 exonic SNPs and genomic
 RT characterization of the IVS12 minisatellite.";
 RL J. Hum. Genet. 48:107-109(2003).
 RN VARIANTS GLN-355; SER-441 AND 857-CYS--THR-858 DEL.
 RX PubMed=15221963; DOI=10.1002/ljc.20269;
 RA Nishijo K., Nakayama T., Aoyama T., Okamoto T., Ishibe T., Yasura K.,
 RA Shima Y., Shibata K.R., Teuboyama T., Nakamura T., Toguchida J.,
 RT "Mutation analysis of the RECQL4 gene in sporadic osteosarcomas.";
 RL Int. J. Cancer 111:367-372(2004).
 RN -1- FUNCTION: DNA-dependent ATPase. May modulate chromosome
 RN segregation.
 CC -1- SUBUNIT: Interacts with UBR1 and UBR2.
 CC -1- SUBCELLULAR LOCATION: Nuclei and cytoplasmic.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed, with highest levels in
 CC thymus and testis.
 CC -1- INDUCTION: Up-regulated in actively dividing cells.
 CC -1- DISEASE: Defects in RECQL4 are a cause of Rothmund-Thomson
 CC syndrome (RTS) [MIM:268400]. A disease characterized by
 CC dermatological features such as atrophy, pigmentation, and
 CC telangiectasia and frequently accompanied by juvenile cataract,
 CC saddle nose, congenital bone defects, disturbances of hair growth,
 CC and hypogonadism.
 CC -1- DISEASE: Defects in RECQL4 are a cause of RAPADILINO syndrome
 CC [MIM:266280]. A disease characterized by radial and patellar
 CC aplasia or hypoplasia.
 CC -1- SIMILARITY: Belongs to the helicase family. RecQ subfamily.
 CC -1- DATABSE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.inbioigen.fr/services/chromocancer/Genes/RECQL4ID285.html".
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL: AB006532; BAA74453.1; -; mRNA.
 CC EMBL: AB026546; BAA6899.1; -; Genomic_DNA.
 CC EMBL: BC011602; AAH11602.2; -; mRNA.
 CC EMBL: BC013277; AAH13277.2; ALT_INIT; mRNA.
 CC HSSP: P15043; LOY.
 CC Ensemble: ENSG00000160957; Homo sapiens.
 CC HGNC: HGNC:9949; RECQL4.

DR MIM; 603780; -;
 DR MIM; 268400; -;
 DR MIM; 266280; -;
 DR GO; GO:0003678; F:DNA helicase activity; TAS.
 DR GO; GO:0007275; P:development; TAS.
 DR GO; GO:0006281; P:DNA repair; TAS.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR01545; DEAD/DEAH_N.
 DR InterPro; IPR002464; DEAH_box.
 DR InterPro; IPR001650; Helicase_C.
 DR InterPro; IPR004589; RecQ.
 DR PANTHER; PTHR13710; RecQ; 1.
 DR Pfam; PF00271; Helicase_C_1.
 DR TIGRFAMs; TIGR00614; recQ_fam; 1.
 DR PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE NEG.
 KW ATP-binding; Catalact; Helicase; Hydrolyase; Nuclear protein;
 KW Nucleotide-binding; Polymorphism.
 NP BIND 502 509
 FT NP BIND 502 509
 FT MOTIF 605 608
 FT NP BIND 267 267
 FT VARIANT 355 355
 FT VARIANT 441 441
 FT VARIANT 857 858
 FT VARIANT 1005 1005
 SQ SEQUENCE 1208 AA; 133077 MW; CB809A765AB48A1 CR664;
 Query Match 100.0%; Score 1617; DB 1; Length 1208;
 Best local similarity 100.0%; Pred. No. 1.9e-125;
 Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QLTVALDMEBEAETLLCTLELPHNWLBELATTTTCRLNCGSPGAOLALHRCPL 60
 DB 907 QLTVALDMEBEAETLLCTLELPHNWLBELATTTTCRLNCGSPGAOLALHRCPL 966
 QY 61 AVCLAQQLPEDPGGSSSVFEDWKLVDMSGWELASVRALCOLQWHEPRTGVRGTGV 120
 DB 967 AVCLAQQLPEDPGGSSSVFEDWKLVDMSGWELASVRALCOLQWHEPRTGVRGTGV 1026
 QY 121 LVFESLAFHLRSPGDLTAEEKQICDPLVGRVAREARLARLRTPQAFHSAFSCG 180
 DB 1027 LVFESLAFHLRSPGDLTAEEKQICDPLVGRVAREARLARLRTPQAFHSAFSCG 1086
 QY 181 PCLFQDDEERSTRKDLGRFYFEEBEGQBGWEDAGPFGQARLDWEDVRCIDRPF 240
 DB 1087 PCLFQDDEERSTRKDLGRFYFEEBEGQBGWEDAGPFGQARLDWEDVRCIDRPF 1146
 QY 241 LSLRPEKFSRAVARIFHGIGSPCYPAOYVGODRRFMRYLHLSFALVGLATEELLQV 300
 DB 1147 LSLRPEKFSRAVARIFHGIGSPCYPAOYVGODRRFMRYLHLSFALVGLATEELLQV 1206
 QY 301 AR 302
 DB 1207 AR 1208
 RESULT 2
 RECO4 MOUSE STANDARD; PRT; 1216 AA.
 AC 075NR7; 076MT1; 099PV9;
 DT 13-SEP-2005 (Rel. 48, Created)
 DT 13-SEP-2005 (Rel. 48, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE ATP-dependent DNA helicase Q4 (EC 3.6.1.-) (RecQ protein-like 4).
 GN Name=RecQ4; Synonyms=Recq4;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

CC Muroidea; Muridae; Murinae; Mus.
 CX NCBI_TaxID=10090;.
 RN [1]
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA / mRNA] (ISOFORM 1) AND NUCLEOTIDE
 RP SEQUENCE [mRNA] OF 16-1115 (ISOFORM 2).
 RX PubMed=1167012;
 RA Ohnata T., Araki R., Fukumura R., Kuroiwa A., Matsuda Y., Tatsuami K.,
 RA Abe M.;
 RT "Cloning, genomic structure and chromosomal localization of the gene
 RT encoding mouse DNA helicase Recq helicase protein-like 4.";
 RL Gene 261:251-258(2000).
 RN [2]
 RP FUNCTION
 RX PubMed=12915449; DOI=10.1093/jmg/dqg254;
 RX Hoki Y., Araki R., Fujimori A., Ohnata T., Koseki H., Fukumura R.,
 RA Nakamura M., Takahashi H., Noda Y., Kito S., Abe M.;
 RT "Growth retardation and skin abnormalities of the Recq4-deficient
 RT mouse";
 RL Hum. Mol. Genet. 12:2293-2299(2003).
 RN [3]
 RP DEVELOPMENTAL STAGE.
 RX PubMed=12952869; DOI=10.1093/jmg/ddg306;
 RA Sittonen H.A., Kopra O., Kaeaeistinen H., Haravuori H., Winter R.M.,
 RA Saarenen A.-M., Peltonen L., Kestinen M., Haravuori H., Winter R.M.,
 RA "Molecular defect of RAPADILINO syndrome expands the phenotype
 RT spectrum of RECQL diseases.";
 RL Hum. Mol. Genet. 12:2837-2844(2003).
 RN [4]
 RP FUNCTION.
 RX PubMed=15703196; DOI=10.1093/jmg/ddi1075;
 RA Mann M.B., Hodges C.A., Barnes E., Vogel T.J., Luo G.;
 RA "Defective sister-chromatid cohesion, aneuploidy and cancer
 RT predisposition in a mouse model of type II Rothmund-Thomson
 RT syndrome";
 RL Hum. Mol. Genet. 14:813-825(2005).
 CC -1- FUNCTION: DNA-dependent ATPase. (By similarity). May play a role in
 CC development of the palate and the limbs. May modulate chromosome
 CC segregation.
 CC -1- SUBUNIT: Interacts with UBR1 and UBR2 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q75NR7-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q75NR7-2; Sequence=VSP_015177;
 CC -1- DEVELOPMENTAL STAGE: Not expressed at E12.5. Expressed at E15.5-
 CC E18.5, with highest levels in chondrocytes of developing bone and
 CC cartilage and immature proliferating enterocytes of intestine.
 CC -1- MISCELLANEOUS: Absence of the protein causes early embryonic
 CC lethality. Transgenic mice with exon 13-deleted RECQL4 are severely
 CC growth-retarded and show high (95%) perinatal lethality. They
 CC exhibit various skin, bone, intestine, tooth and thymus
 CC abnormalities and premature aging features, but have normal
 CC sensitivity to IR and UV irradiation. In contrast, transgenic mice
 CC expressing a truncated form of RECQL4 exhibit mild perinatal
 CC lethality, no growth defect, but show defects of the skin and
 CC skeleton, aneuploidy and increased cancer susceptibility.
 CC -1- SIMILARITY: Belongs to the helicase family. Recq subfamily.
 CC -1- SIMILARITY: Contains 1 CCHC-type zinc finger.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; AB039882; BAD1131.1; -; mRNA.
 CC EMBL; AB175741; BAD14289.1; -; mRNA.
 CC EMBL; AB042529; BAB32696.1; -; Genomic_DNA.
 CC HSP; P15043; 10Y.
 CC Ensembl; ENSMUSG0000003762; Mus musculus.
 CC GGI; GGI:1931028; Recq14.

DR PROSITE PS00690; DEAH ATP HELICASE; FALSE_NEG.
 DR PROSITE PSS0158; ZF_CCHC_1.
 KM Alternative splicing; ATP-binding; Helicase; Hydrolase; Metal-binding;
 KW Nuclear protein; Nucleotide-binding; Zinc; Zinc-finger.
 FT ZN_FING 393 410 CCHC-type.
 FT NP_BIND 519 526 ATP (Potential).
 FT MOTIF 627 630 DEAH box.
 FT VARSPIC 843 843 O -> OVGSFGPSIPQDPRSTTIRPPOPOLSLCLPVSCRP
 GPGKSGVHTVMQ (in isoform 2).
 /FTID=VSP_015177.
 ERIQNGWL -> RKDTERRVA (in Ref. 1;
 BAD1131).

Query Match 70.8%; Score 1144.5; DB 1; Length 1216;
 Best Local Similarity 73.2%; Pred. No. 3.9e-86;
 Matches 218; Conservative 26; Mismatches 53; Indels 1; Gaps 1;

OY	1	QLTVQALDMBEAEIETLLCYLELHPHMLBLATYYTHCRINCPGSPAOLALAHRCPL	60
DB	916	QSTVALDMTEBAEITLLCTYLELHPRMLELPPTYAQCHLHCIGSAGLOALAHRCPL	975
OY	61	AVLCIAQGLPEDPGSSSVPEFDWKLVDNSGMELASVRRLCOLQMDHEPTGYRRTGV	120
DB	976	AACAKMPKPQTSGRSLSLEGVELLDNSMGWKLASVRQALHLQKWPEPKKGAAOGGV	1035
OY	121	LVSSELAFLHSRGDLTAERKDCIPELVGRVQAERQALARLRFTGAHFSAFPSCG	180
DB	1036	LVKFSFLAFLHSRGDLTDDEKKDCIPELVNRVQAREKKAHLHLQMSKAFRSAVPSCG	1095
OY	181	PCLSGODEBERTRKDLIGRFEEBEGCEPCGMEDAQGEPGQARLODMEOVCNTIQF	240
DB	1096	PCLDSNEBHNSNQKTLVSYTFEEB-EFEETMTDITDGPKKGQVLQDMENQIRDVQL	1155
OY	241	LSLRPEEFKFSRAVARIFPHGISGPCYPAYVGODRRFMFKYLHLSPHALVGLATEELL	298
DB	1155	LSLRPEERFSGRAVARIFPHGISPCYPAYVGGLDRFRFKYLHLDFHALMILATEELL	1212

RESULT 3
 QJUNX8 XENLA PRELIMINARY; PRT; 1500 AA.

ID	QJUNX8_XENLA	PRELIMINARY; PRT; 1500 AA.
AC	QJUNX8_	
DT	13-SEP-2005 (TREMBLrel. 31, Created)	
DI	13-SEP-2005 (TREMBLrel. 31, Last sequence update)	
DR	13-SEP-2005 (TREMBLrel. 31, Last annotation update)	
DE	RECOLD4-helicase-like protein.	
CN	Name=RFS;	
OS	Xenopus laevis (African clawed frog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;	
CC	Xenopodinae; Xenopus; Xenopus.	
OX	NCBI_Taxid=8355;	
RX	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RA	Sangrithi M.N., Bernal J.A., Madine M., Philpott A., Lee J.,	
RT	Dunphy W.G., Venkataraman A.R.;	
RT	"Initiation of DNA replication requires the RECOLD4 protein mutated in	
RU	Rotmund-Thomson syndrome.";	
RU	Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.	
DR	EMBL, DQ059311; AAY89585.1; -, mRNA.	
KW	Helicase.	
SQ	SEQUENCE 1500 AA; 168706 MW; A9SF0A2BB2BE554D CRC64;	

Query Match 51.5%; Score 833; DB 2; Length 1500;
 Best Local Similarity 54.6%; Pred. No. 4e-60;
 Matches 165; Conservative 53; Mismatches 78; Indels 6; Gaps 4

3 TVQALDMBEAEIETLLCYLELHPHMLBLATYYTHCRINCPGSPAOLALAHRCPLAV 62

```
Db 1200 TVGSLDIREAIEITMLCYLEINSPHMLFQHLPTLSHCIVVCSGQQLRLARSPFPAV 1259
Qy 63 CLA-QQLPEDPGQSSSVFEDMVKLVDSMGWELASVRRALCOLQMDHEPRTGVRGT--- 118
Db 1260 CLARERLGVGHTHTVSSSTFENVVELADSMGWEVAVVKALNDLQVRBPDKGF-KGTGRS 1318
Qy 119 GVLFPSRLAFHLSRPGDLTAEEKQICDPLXGRVQAEERQALARTPTQAFHSVAPPS 178
Db 1319 GVLFPSRLAFHLSRPGDLTAEEKQICDPLXGRVQAEERQALARTPTQAFHSVAPPS 1378
Qy 179 CGPCLGQDEBERSTRKLDLGRYFEESGQEPGEMEDAQGPPEQARLQMDQVRCDCIR 238
Db 1379 SSACAEKMSARSRASQKLKLLDYFEKRDTELEAKVEDD-BELQIKYQEAQDVRSVDR 1437
Qy 239 QPLSLRPEKFSRAVARIFHGIGSPCYPAQVYGDRFRFWRKYLHLSFHALVGLATEEL 298
Db 1438 HFLSIHEBERPSGRALARIIFHGIGSPCYPAQVYGRDRFRFWRKYLHLSFHALVGLATEEL 1497
Qy 299 QY 300
Db 1498 RL 1499

RESULT 4
Q4RLC3_TETNG PRELIMINARY; PRT; 910 AA.
ID Q4RLC3_TETNG PRELIMINARY; PRT; 910 AA.
AC Q4RLC3;
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Chromosome 21 SCAF15022, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0032550001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neocentropomidae;
OC Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OC NCBI_TaxID=99883;
RN RP
RP NCBIOTIDE SEQUENCE.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mucelli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicoud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segutrens B.,
RA Da Silva C., Salenouat M., Levy M., Boudet N., Castellano S.,
RA Blemond C., Jubin C., Castell V., Kacinka M., Vacherie B.,
RA Blemond C., Skalli Z., Cactolico L., Poulain J., De Bernardis V.,
RA Craud C., Duprat S., Broctier P., Coutanceau J.P., Gouy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan F., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Lander V., Schachter V., Queller F., Saurin W., Scarpelli C.,
RA Wincker P., Landier B.S., Weissenbach J., Roest Crolious H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN RP
RP NCBIOTIDE SEQUENCE.
RG Genoscope, Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; CAAB01015022; CAG10809.1; -; Genomic_DNA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; zf-CCHC; 1.
```

```
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00467; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00343; ZNF_C2HC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS0158; 2P_CCHC; 1.
KW ATP-binding; Helicase; Hydrolase.
FT NON TER
FT NON TER
SQ SEQUENCE 910 AA; 101865 MW; 7CEFF693FB4564F CRC64;

Query Match 46.6%; Score 753.5; DB 2; Length 910;
Best Local Similarity 50.2%; Pred. No. 9.2e-54;
Matches 151; Conservative 56; Mismatches 85; Indels 9; Gaps 5;

Qy 1 QLTVALDMPREAEITLLCYLELPHHMLBILATTTYTHCRNLCPGPAQALAHRCPL 60
Db 615 QQTVEALDITBEGVETILLCYELHLPQRFVBLHPTLSVCKISCYDGRQLRRVTKICPV 674
Qy 61 AVCLIA-QQLPEDPGQSSSVFEDMVKLVDSMGWELASVRRALCOLQMDHEPRTGVRG-T 118
Db 675 AVVLARRRMAGERVESCDQLEFVVEVADTWGMQLPLVKGLROLQWS---TGQGGRS 730
Qy 119 GVLFPSRLAFHLSRPGDLTAEEKQICDPLXGRVQAEERQALARTPTQAFHSVAPPS 178
Db 731 GVLFPSRLAFHLSRPGDLTAEEKQICDPLXGRVQAEERQALARTPTQAFHSVAPPS 790
Qy 179 CGPCLGQDEBERSTRKLDLGRYFEER-BEQEPGEMEDAQGPPEQARLQMDQVRCDCI 237
Db 791 VSSGQDLDLPTDRSLQKLSLSSEYFDKRRBDRAVPAPEDYGLD--RKVLDMEQITADI 848
Qy 238 RQFLSLRPEKFSRAVARIFHGIGSPCYPAQVYGDRFRFWRKYLHLSFHALVGLATEEL 297
Db 849 RSFLANRSDEKFSRAVARILHIGISPCYPAQTYGKDRFRFWRKYLHLSFHALVGLATEEL 908
Qy 298 L 298
Db 909 I 909

RESULT 5
Q7PNM8_ANOGA PRELIMINARY; PRT; 1058 AA.
ID Q7PNM8_ANOGA PRELIMINARY; PRT; 1058 AA.
AC Q7PNM8;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE ENSANGP00000004546 (Fragment).
GN ORFNames=BNSANG50000003572;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;
OC Anophelinae; Anopheles.
OC NCBI_TaxID=180454;
RN RP
RP NCBIOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN RP
RP NCBIOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAAB01008960; EAAL1913.3; -; Genomic_DNA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0016787; F:Hydrolase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
```

DR GO:0006259; P:DNA metabolism; IEA.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH_N.
DR InterPro: IPR01650; Helicase_C.
DR InterPro: IPR004589; RecQ.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00098; zf_CCHC; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
DR TIGRPFAMs: TIGR00614; recQ_fam; 1.
DR PROSITE: PS50158; ZF_CCHC; 1.
DR ATP-binding; Helicase; Hydrolase.
KW ATP-binding; Helicase; Hydrolase.
FT NON TER 1
SQ SEQUENCE 1058 AA; 117711 MW; 6DPBPF17EA2255FC CRC64;

Query Match 28.7%; Score 464.5; DB 2; Length 1058;
Best Local Similarity 36.3%; Pred. No. 1.2e-29;
Matches 110; Conservative 56; Mismatches 122; Indels 15; Gaps 7;

QY 1 QLTVALDMPBEAIEITLTCYLELHPHMLBLATTYTHCRINCPGPAQLALAHRCPL 60
DB 766 EATVQQLDIPENITTFICYELDEQRITQALSPAYTMCKWMSYGVAPLFGAAKECPPL 825
QY 61 AVCLAQQLPEDPGQ-SSSVPEFDMVKLVDSMGWELASVRALCOLQMDHEPRTGRTG 119
DB 826 AMAFLDLKRGKISHSTAIETFPVIDVASAIGMDGVVYQLKNLEM--TTVNNVRKRSP 883
QY 120 VLVESELAFLRSRPGDITAEKQICDPLYGVRQARERQALRLRRTFOAFHSAFSPSC 179
DB 884 LSVFTTELGRFRARAGDLTDELDHALDGLVRYVHQRTQALQYISDALNSVCENTT 943
QY 180 GPCEEQD--EERSYRLKDLGRYFEEBEGQPGMEDAQGPQARLQMDQVRCDI 237
DB 944 GP-VERACCPGSPDKLTIYREYFTTDSKE---QLEIVPEPDDT---DEGLINDI 994
QY 238 RQFSLRPEKFPSSAVARIFHGIGSPCYPAQVYQDRFWRKYLLHSFHALVGLATBEL 297
DB 995 RTALCRYPENFTGAIARLPHGVQSPYVALVMSRS-NFWRAVYKTDPNRIVRLANAEI 1053
QY 298 LQV 300
DB 1054 VRM 1056

RESULT 6

Q9NH11_DROME PRELIMINARY; PRT; 1530 AA.
AC Q9NH11;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE RECO4.
GN Name=RecQ4; Synonyms=RECQ4; ORFNames=CG7487;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kusano K., Bettes W.B., Engels W.R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF236559; AAF42939.1; -; Genomic_DNA.
DR HSSP; P15043; 10YV.
DR Intact; Q9NH11; -;
DR FlyBase; FBgn0040290; CG7487.
DR FlyBase; FBgn0040290; RecQ4.
DR GO:0005524; F:ATP binding; IEA.
DR GO:0008026; F:ATP-dependent helicase activity; IEA.

DR GO:0003676; F:nucleic acid binding; IEA.
DR GO:0006259; P:DNA metabolism; IEA.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR011545; DEAD/DEAH_N.
DR InterPro: IPR01650; Helicase_C.
DR InterPro: IPR004589; RecQ.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF00271; Helicase_C; 1.
DR Pfam: PF00098; zf_CCHC; 1.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00343; Znf_C2HC; 1.
DR TIGRPFAMs: TIGR00614; recQ_fam; 1.
DR PROSITE: PS50158; ZF_CCHC; 1.
SQ SEQUENCE 1530 AA; 169878 MW; 1B753A59A3E30EB4 CRC64;

Query Match 26.4%; Score 427; DB 2; Length 1530;
Best Local Similarity 31.1%; Pred. No. 2.4e-26;
Matches 95; Conservative 62; Mismatches 122; Indels 26; Gaps 6;

QY 3 TVQLDMPBEAIEITLTCYLELHPHMLBLATTYTHCRINCPGPAQLALAHRCPLAV 62
DB 1241 TVEMLDIPAEINISTLTCTMEIDPRWCISVASAVMAKVSYGPKYLKHAKECPPLAM 1300
QY 63 CLAQQLPEDP-GQSSSVPEFDMVKLVDSMGWELASVRALCOLQMDHEPRTGRTGVL 121
DB 1301 ALAQLIRDKRTKEQSNITIEFSDITDAIGIGNSGVVKQLKDLN--VKVNGYRKRSPT 1358
QY 122 VEFSELAFLRSRPGDITAEKQICDPLYGVRQARERQALRLRRTFOAFHSAFSPSCP 181
DB 1359 VSFYDLGRIRIVPDDFTESEIDNALDITLTSVQGERQLQLQVANGLAAVASSGQ 1418
QY 182 CLEQD-DEERSTRKDLGRYFEEBEGQ-----EPGMEDAQGPQARLQMDQVRC 234
DB 1419 CCNADFQDRGEOLKATVRNFANDYPQDLELEIEPSPVDP-----ENIT 1463
QY 235 CDIQFSLRPEKFPSSAVARIFHGIGSPCYPAQVYQDRFWRKYLLHSFHALVGLAT 294
DB 1464 DVHALLIMYPDNTGTGNIRARIFHGIGSPYPAVITWR-CKFRAHVAKVDPNRLHLAN 1522
QY 295 BELQ 299
DB 1523 MAIK 1527

RESULT 7

Q9VSE6_DROME PRELIMINARY; PRT; 1579 AA.
AC Q9VSE6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE CG7487-PA.
GN Name=RecQ4; ORFNames=CG7487;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Menthlov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Ceiniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Paley J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminier U.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Ceiniker S.E.,
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tuzy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Ceiniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.,
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Ceiniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Paley J., Park S., Svirskas R., Smith E.,
RA Yu C., Rubin G.,
RT "Drosophila melanogaster release 4 sequence."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP NUCLEOTIDE SEQUENCE.

RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003556; AAF50477.1; -, Genomic_DNA.
DR HSSP; P15043; 10Y.
DR Ensemble; CG7487; Drosophila melanogaster.
DR FlyBase; FBgn0040290; CG7487.
DR FlyBase; FBgn0040290; RecQ4.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00339; C2HCZNFINGER.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00343; ZNF_C2HC; 1.
DR TIGRFAMs; TIGR00614; recQ fam; 1.
DR PROSITE; PS50158; ZF_CCHC_1.
SQ SEQUENCE 1579 AA; 175909 MW; EA24B9B49C8FCF7D CRC64;

Query Match 26.4%; Score 427; DB 2; Length 1579;
Best Local Similarity 31.1%; Pred. No. 2.5e-26;
Matches 95; Conservative 62; Mismatches 122; Indels 26; Gaps 6;

QY 3 TVQALDMPBEAIEITLTCYELHPPHMLEATTYTHRLNCPGPAOLQALAHRCPLAV 62
DB 1290 TVEMLDIPAEIVITLTCYMEIDPRWCISVLSAYVAKVISYGGPKYIKNAKXCPPLAM 1349
QY 63 CLAQLLEDP- GQSSSVFPMVVLVDSMGMELASVRALCOLQMDHEPRTGVARGGYL 121
DB 1350 AIAQLQDKTKFKESNSITFSVDIAAGIGNSGVVYKLDLEW--VKVNGYKRSPTT 1407
QY 122 VEFSELAFLHRLSPGDLTAEKQICDPLVGRVQARROALARLRTQAHSVAFSPGCP 181
DB 1408 VSFYDLGFRKVPDPDFESEIDNALDITLTVRSVQERTQLIQYVHAGLAVAAYSSCGQ 1467
QY 182 CLEEQ-DEBSTRKLDLGRFYEEBEGQ-----EPGMEDAQBPBGQARLQDWEQVR 234
DB 1468 CCNADFPQDRGEQLKAVRVYFANDYPQDLELEIEPSNVDP-----ENII 1512
QY 235 CDIRQFSLRPEBEKFSRAVARIFHGIGSPCYPAQVYGDRRFMRKYLHLSFHALVGLAT 294
DB 1513 DDVALINMTPDNTFTGRNIALIFHGIMSPVPAVIMGR-CKFWRAHVKVDFRILHLAN 1571
QY 295 EELQ 299
DB 1572 MAIK 1576

RESULT 8
ID Q9C6N0 ARATH PRELIMINARY; PRT; 941 AA.
AC Q9C6N0.
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ATP-dependent DNA helicase RecQ, putative.
GN Name=F28L5.4;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,

RA Maiti R., Romning C.M., Koo H., Fujii C.Y., Uterback T.R.,
RA Bainsread M.E., Bowman C.L., White O., Niernan W.C., Frazer C.M.;
RU Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
CC -I- FUNCTION: EIF4A is both a subunit of a high molecular weight
CC protein complex involved in cap recognition and is required as a
CC single polypeptide chain for mRNA binding to ribosome. It is an
CC ATP-dependent single stranded DNA-binding protein with a sequence-
CC independent unwinding activity (helicase) (By similarity).
CC -I- SUBUNIT: eIF4F is a multi-subunit complex, the composition of
CC which varies with external and internal environmental conditions.
CC It is composed of at least EIF4A, EIF4B and EIF4G (By similarity).
CC EMBL; AC079280; AAG50580.1; -; Genomic_DNA.
DR HSSP; P15043; 10YX.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002464; DEAH_Box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS00690; DEAH ATP HELICASE; UNKNOWN_1.
DR ATP-binding; Helicase; Hydrolyase; Initiation factor; Nuclear protein;
KW Protein biosynthesis.
SQ SEQUENCE 941 AA; 104156 MW; 30BC454C795522E1 CRC64;

Query Match 11.5%; Score 186; DB 2; Length 941;

Best Local Similarity 24.4%; Pred. No. 1.5e-06;
Matches 77; Conservative 45; Mismatches 126; Indels 66; Gaps 12;

QY 5 QALDMPERAIETLLCYLELPHHMLLATTYTHCRINCPCGPAQLALAHRCPLA 61
DB 659 QKFDKKEEVMQTLTLHLELGEVYLRLPQLNICTLN-----FHKTSDAQNIV 707
QY 62 VCLAQQLPEDPGCGSSSVFEDMKLVDSMG--WELASPRALCOLQWHEPRGVRRCT 118
DB 708 MLYVGSPPMTLARSATVAAILKKSHVKGQGLHVFDPVAVASSICVAT-----T 755
QY 119 GVLVSESELAFHLRSPGDLTAEEKD-----QICDFLYGRVQARERQALALRR 166
DB 756 DVLAIEIQL-----KGEVYTELKDSAFCTYTIKSKKEICLSLHILTKMLTEIESCKYAK 809
QY 167 -TFQAFHSVAFSPGCPCLQEODEERSTRKDLGRYFEEBEGQEPGMEDAQEPFGQAR 225
DB 810 LDIMSSAAVAISVNTSELSSGAKQTR--SLQSRIFDYFNG-----DEKDSPSKA- 859
QY 226 LQDMEQD---VRCIDRQPLSLRPEKSSRAVAIFHGIGSPCYPAQYGGDRFRWKY 281
DB 860 -----TONCAFIRADIKVFLQSNRQAKFTPRALAIRIMHGVSAPFNSVWSK-THFWGRY 913
QY 282 LHLSPHALVGLATEEL 297
DB 914 MNVDPRVIMEAQTEL 929

RESULT 9

Q8W028_ARATH PRELIMINARY; PRT; 870 AA.
AC Q8W028;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Helicase.

GN Name=rgl-5;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OK NCBI_TaxID=3702;

RP NUCLEOTIDE SEQUENCE.

RA Hartung F., Plachova H., Puchta H.;
RU Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.

CC -I- FUNCTION: EIF4A is both a subunit of a high molecular weight
CC protein complex involved in cap recognition and is required as a
CC single polypeptide chain for mRNA binding to ribosome. It is an
CC ATP-dependent single stranded DNA-binding protein with a sequence-
CC independent unwinding activity (helicase) (By similarity).
CC -I- SUBUNIT: eIF4F is a multi-subunit complex, the composition of
CC which varies with external and internal environmental conditions.
CC It is composed of at least EIF4A, EIF4B and EIF4G (By similarity).
CC EMBL; AJ421618; CAD13472.1; -; mRNA.
DR HSSP; P15043; 10YX.

DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.

DR InterPro; IPR001410; DEAD.
DR InterPro; IPR011545; DEAD/DEAH_N.
DR InterPro; IPR002464; DEAH_Box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.

DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.

DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.

DR TIGRFAMs; TIGR00614; recQ_fam; 1.
DR PROSITE; PS00690; DEAH ATP HELICASE; UNKNOWN_1.

DR ATP-binding; Helicase; Hydrolyase; Initiation factor; Nuclear protein;
KW Protein biosynthesis.
SQ SEQUENCE 870 AA; 96383 MW; 8B5ADA4F47A81723 CRC64;

Query Match 11.4%; Score 184.5; DB 2; Length 870;

Best Local Similarity 24.3%; Pred. No. 1.9e-06;
Matches 76; Conservative 45; Mismatches 119; Indels 73; Gaps 12;

QY 5 QALDMPERAIETLLCYLELPHHMLLATTYTHCRINCPCGPAQLALAHRCPLAVCL 64
DB 599 QKFDKKEEVMQTLTLHLELGEVYLRLPQLNICTLN-----FHKSSPNTLAA 647
QY 65 AQQLPEDPGCGSSSVFEDMKLVDSMG--WELASPRALCOLQWHEPRGVRRCTGL 121
DB 648 R-----SAYVAAILKKSHVKGQGLHVFDPVAVASSICVAT-----TDVL 685
QY 122 VEFSELAFLHRSBGDLTAEEKD-----QICDFLYGRVQARERQALALRR--TF 168
DB 686 AEIQ-----LKKKGEVTELEKDSAFCTYTIKSKKEICLSLHILTKMLTEIESCKRKLDI 741
QY 169 QAFHSVAFSPGCPCLQEODEERSTRKDLGRYFEEBEGQEPGMEDAQEPFGQARLQD 228
DB 742 MSSAAVAISVNTSELSSGAKQTR--SLQSRIFDYFNG-----DEKDSPSKA---- 788
QY 229 WEDQ---VRCIDRQPLSLRPEKSSRAVAIFHGIGSPCYPAQYGGDRFRWKYHL 284
DB 789 --TONCAFIRADIKVFLQSNRQAKFTPRALAIRIMHGVSAPFNSVWSK-THFWGRYMN 845
QY 285 SFHALVGLATEEL 297
DB 846 DFRVIMEAQTEL 858

RESULT 10

[illegible]

Dd Db 791 ----- RTLOQKSSPFLQADIVFIQSNSFAFTPPRAVARIMHGISSPAFVTSWKN 842

Oy DB 275 RRFWRKYLLSHFALVGATLELLOW 300
 || :| : | :|| :| :
Db 843 -HFMGRYEVDFPLVMBAKKELVKL 867

RESULT 11
O7X829 ORYSA
ID O7X829 ORYSA PRELIMINARY; PRT; 927 AA.
AC Q7X829
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DE 01-MAR-2004 (TREMblrel. 26, last annotation update)
GN OJ99113_30.2 protein.
NM Name=Oj99113_30.2;
OS Oriza sativa [Japanica cultivar-group].
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyte; Magnoliophyta; Liliopsida; Poales; Poaceae;
OX Ehrhartidiaceae; Oryzaceae; Oryza.
NCBI_TaxId=39947;

[1]
NP NUCLEOTIDE SEQUENCE
RX MEDLINE=2237377; PubMed=12447433; DOI=10.1038/nature01183;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Weng Q., Zhang L., Lu Y., Mu T., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang K., Yin H.,
CAI Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
HA Han B.;
RL "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
CC CC -1- FUNCTION: EIF4A is both a subunit of a high molecular weight
protein complex involved in cap recognition and is required as a
single polypeptide chain for mRNA binding to ribosome. It is an
ATP-dependent single stranded DNA-binding protein with a seque-
CC CC nce-independent unwinding activity (helicase) (By similarity).
CC CC -1- SUBUNIT: eIF4F is a multi-subunit complex, the composition of
which varies with external and internal environmental conditions.
CC CC It is composed of at least EIF4A, EIF4B and EIF4G (By similarity).
DR DR EMBL: AL662946; CAD11320.2; Genomic_DNA.
DR DR HSSP; PL5043; IOXY.

DR GRameine; Q7X829; --
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008026; F:ATP-dependent helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003743; F:translation initiation factor activity; IEA.
DR GO; GO:0006259; P:DNA metabolism; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001410; DEAd.
DR InterPro; IPR011545; DEAD/DEAF_N.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR004589; RecQ.
DR Pfam; PF00270; DEAD_1.
DR Pfam; PF00271; Helicase_C_1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_G; 1.
DR TIGRfams; TIGR00614; recQ_fam; 1.
KW ATP-binding; Helicase; Hydrolase; Initiation factor;
KW Protein biosynthesis.

SQ SEQUENCE 927 AA; 102031 MW; 4AB7BC5504C04FFA CRC64;

Query Match 10.5%; Score 169; DB 2; Length 927;
Best Local Similarity 20.9%; Pred.No. 3.9e-05;
Matches 68; Conservative 58; Mismatches 118; Indels 82; Gaps 12;

1 QLTVQADMPEAEAYETILLCYEHLPHNHMLBLATTYYTHCRINCGSPQAOLALAHRCPLL 60

```
DB 651 ELTSKPKDKEVLTITLTQIEGQOQIRLLPQSVCTL-----YFHKTSPO 699
OY 61 AVC-----LAQQLPEPQGGSSSVEPDMKYLDMSKMEELASVRALCOLQMDHEPTGYR 116
DB 700 LLADKDLIRSLNRSNEMKDGHYVPDIPRIANDLKITNNEV-----FHLNR----- 746
OY 117 GGVGLVRS-ELAFHLRSPG-----DLTAEKQICDPLGRVQAEERQALRAR 165
DB 747 -----LKSSELSFELKDPAYCYVILMPDDNALSN-----LTKMSSEVSSKISLSD 796
OY 166 RTFOAFHSVAPFSC-----GPCLEQDEERSTRLKDLGRYEEERGQEPGME 214
DB 797 AMF-ALANPAPVAGCKRTGGCGSGSGHTPCIOKKIME-----YFSKDDGTSENDC- 843
OY 215 DAQGEPEQARLQDMEVDVRCIRPFLSRPEKSSRAVANIFGIGSPCEPAQVYQOD 274
DB 844 -----RTQLQKSSPFLQADIKVFIOGNSPAPKFTPRAVARIMGIGSSPAPSVTWSKN 895
OY 275 RRFMRKYLHSPHALVGLATEBELQV 300
DB 896 -HFMGRVVEVDPLVMEAKAKELVXL 920

RESULT 12
TRIS6_HUMAN
ID TRIS6_HUMAN STANDARD; PRT: 755 AA.
AC Q9BRZ2; Q9BRV7; Q9NZH8; Q9NAC0; Q9H031;
DT 01-FEB-2005 (Rel. 46, Created)
DT 10-FEB-2005 (Rel. 46, Last sequence update)
DT 01-MAY-2005 (Rel. 47, Last annotation update)
DE Tripartite motif protein 56 (RING finger protein 109).
GN Name=TRIS6; Synonyms=RNP109;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homidae;
OC Homio.
NCBI_TaxID=9606;

[1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP MEDLINE=22737999; PubMed=12853948; DOI=10.1038/nature01782;
RX Hillier L.W., Fulton R.S., Fulton L.A., Graves T.A., Pepin K.H.,
Wagner-McPherson C., Layman D., Maas J., Jaeger S., Walker R.,
Wyllie K., Sekhon M., Becker M.C., O'Laughlin M.D., Schaller M.E.,
Ruehl G.A., Delenau K.D., Miner T.L., Nash W.E., Cordes M., Du H.,
Sun H., Edwards J., Bradshaw-Cordum H., Ali J., Andrews S., Isak A.,
Vandrun A., Nguyen C., Du F., Lamar B., Courtney L., Kalicki J.,
Ozerky P., Bielicki L., Scott K., Holmes A., Harkins R., Harris A.,
Strong C.M., Hou S., Tomlinson C., Dauphin-Kohlberg S.,
Kozlovicz-Reilly A., Leonard S., Rohlfing T., Rock S.M.,
Tin-Woliam A.-M., Abbott A., Minx P., Maupin R., Strowatt C.,
Latreille P., Miller N., Johnson D., Murray J.W., Woessner J.P.,
Wend M.C., Yang S.-P., Schultz B.R., Wallis J.W., Splich J.,
Bieri T.A., Nelson J.O., Berkowicz N., Wohldmann P.E., Cook L.L.,
Hickenholtz M.T., Eldred J., Williams D., Bedell J.A., Mardis E.R.,
Clifton S.W., Chisoe S.L., Marra M.A., Raymond C., Haugen E.,
Gillett W., Zhou Y., James R., Phelps K., Iadonco S., Buty K.,
Simme E., Levy R., Clendinning J., Kaul R., Kent W.J., Furey T.S.,
Baerisch R.A., Brent M.R., Keibler E., Flicek P., Bork P., Suyama M.,
Bailey J.A., Portnoy M.E., Torbett D., Chinwalla A.T., Glis W.R.,
Eddy S.R., McPherson J.D., Olson M.V., Eichler E.E., Green E.D.,
Waterson R.H., Wilson R.K.,
"the DNA sequence of human chromosome 7.",
Nature 424:157-164(2003).
RL
RN
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Lymph, and Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner C.M., Shenmen G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
Hopkins R.F., Jordan H., Moore T., Max S.J., Wang J., Heit F.,
Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
```

```
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carinici P., Pirange C.,
RA Baka S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Roak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerker A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 1-265 AND 429-755.
RC TISSUE=Spleen, and Thymoid.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Makamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimeta M., Watanabe M., Hirao K., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiani S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Munesashio K., Yuuki F., Oshima A., Sasaki N., Aotaka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RL cDNAs."
Nat. Genet. 36:40-45(2004).
RN
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 572-755.
RC TISSUE=Lymph node;
RG The German cDNA consortium;
RX Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RL
CC 1- SIMILARITY: Belongs to the TRIM/RBCC family.
CC 1- SIMILARITY: Contains 2 B box-type zinc fingers.
CC 1- SIMILARITY: Contains 1 RING-type zinc finger.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL; BC005847; AAH05847.2; ALT_INIT; mRNA.
CC EMBL; BC048194; AAH048194.1; -; mRNA.
CC EMBL; AK092927; BAC04004.1; -; mRNA.
CC EMBL; AK075255; BAC11500.1; ALT_INIT; mRNA.
CC EMBL; AL512757; CAC21676.1; -; mRNA.
CC HSSP; P29590; 1B0K.
CC Ensemble; ENSG00000169871; Homo sapiens.
CC HGNC; HGNC:19028; TRIM56.
```

DR InterPro: IPR000315; Znf Bbox.
DR InterPro: IPR001841; Znf_RING.
DR Pfam: PF00643; Zf-B_box_2.
DR Pfam: PF00097; Zf-C3HC4_1.
DR PRINTS: PRO1406; BBOXZNFINGER.
DR SMART: SM00336; BBOX; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00119; ZF_BBOX; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR PROSITE: PS00089; ZF_RING_2; 1.
DR Coiled coil; Metal-binding; Repeat; Zinc; Zinc-finger.
KW Coiled coil; Metal-binding; Repeat; Zinc; Zinc-finger.
FT ZN_FING 21 60 RING-type.
FT ZN_FING 98 149 B box-type 1.
FT ZN_FING 164 205 B box-type 2.
FT COILED 216 314 Potential.
FT COMPBIAS 302 309 Poly-Ala.
FT COMPBIAS 459 464 Poly-Lys.
FT CONFLICT 230 230 L->P (in Ref. 3; BAC04004).
FT CONFLICT 259 269 AEGVLRALLAQ -> CLARTESCKAE (in Ref. 3; BAC04004).
SQ SEQUENCE 755 AA; 81488 MW; CF72D0C8EC9F69E7 CRC64;

Query Match 6.6%; Score 107; DB 1; Length 755;
Best Local Similarity 27.0%; Pred. No. 4.4;
Matches 70; Conservative 23; Mismatches 106; Indels 60; Gaps 10;

QY 11 BEALETLLCYELHPHMLLELATTT-----KRLNCPGPAQLQALHRCPLAVCLA 65
DB 159 BEAERERQAACQPHGEARLFLCOPCSQLCRCKLD-----PHLDHCLPLAEAVR 210
QY 66 QQLPEDEPGQSSSVFPEKLVDSMGWELASVRR-----ALCOLQWD-----HEPRTG 113
DB 211 ABRRLGSG-----LLAGVDNNVLEEARVKEALRLBQARVGTVEEAAG 261
QY 114 VRGTGVLVEFSELAFLHRSFGDLTAERKQICDPLYGRVQ-AERQALAR-----163
DB 262 VLR--ALLAQKQEVYGLRAHVEAAEERARBLAELREGVQVAAFAAFARVLSTGREA 319
QY 164 -----LARTQAHSVAFPSCGPGLEEGODEBSRLK---DLGRVFEESGGEPG 211
DB 320 EILSLBGLAIAQRRLQOLGCPWAPGPAPCLLPQLSLHPGLLDKNCHLRLSFEEOQPOPDG 379
QY 212 GMEBA--QGPEPGQARLD 228
DB 380 GKQDAGTGGEGESQSRRED 398

RESULT 13

Q96D36 HUMAN
ID Q96D36 HUMAN PRELIMINARY; PRT; 809 AA.
AC Q96D36
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DKFZPS86J0619 protein (Fragment).
GN Name=DKFZPS86J0619;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiinae;
OC Homo.
OX NCBI_TaxID=9606;
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=lymph;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,

RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Huilk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywnicki M.I., Skalka U., Smailus D.E.,
RA Scherren A., Schein J.E., Jones S.U.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=lymph;
RA Strausberg R.,
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013367; AAH13367.2; mRNA.
FT NON TER 1
SQ SEQUENCE 809 AA; 89682 MW; 0D57E2040458DB4 CRC64;

Query Match 6.6%; Score 107; DB 2; Length 809;
Best Local Similarity 26.2%; Pred. No. 4.8;
Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;

QY 29 LELATTTTHCRANCPGPAQLQALHRCPLAVCLAQQL-----PEDEPGQS--SSV 79
DB 25 LQALATV-----LSSPPGALVMSM-HRSHFLCPLRLQCOYQRCVPODTGSSFLKV 78
QY 80 EFDVNVKLVDMSGWELASVRRALCOL--QMDHPRGTGVRGTGLVSESLAFH-----130
DB 79 LLOMTQMLDPSGVBGGELRQLMLASQASAGRLSVRG-GILRLAEALAFQDLEVS 137
QY 131 -----LRSPGDLTAERKQICDPLYGRVQAREQALRLRRTFOAFHSA--FPSC 179
DB 138 STVRAVIATLRS-GEQGSVPEPLISKVLQGLIEVRSFHELLTAPFSAADADAAPPAC 196
QY 180 GPCLIEQODEBSRLKDLGRYFEER--GEQR--GMEBA--QGPEPGQARLDQWEDQV 233
DB 197 KPVVV-----VSSL--LQEEPLAGGRGADGSLAVRUGPSSGLLVDMLEML 244
QY 234 RCDIRQPLSLRPREK-----FSSRAVARIFHGIGSPCYPAQVYGDRFMWKY-----281
DB 245 DPEV--VSSCPDQLRLFSRR-----KKGQAQVPS-----FRPYLLTLFTHQ 286
QY 282 -----LHLSFHALVGLATER 296
DB 287 SSWPTLHQICIRVLKGSREQ 306

RESULT 14

Q6NT70 HUMAN
ID Q6NT70 HUMAN PRELIMINARY; PRT; 900 AA.
AC Q6NT70
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE DKFZPS86J0619 protein (Fragment).
GN Name=DKFZPS86J0619;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiinae;
OC Homo.
OX NCBI_TaxID=9606;
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Brain;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:43:16 ; Search time 13 Seconds
(without alignments)
1920.620 Million cell updates/sec

Title: US-09-889-325-4_COPY_907_1208

Perfect score: 1617
Sequence: 1 QLTVALDMPREAIETLLCY.....HLSPHALVGLATELLQVAR 302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents AA:*
2: /cgn2_6/prodata/1/1aa/5 COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6 COMB.pep:*
4: /cgn2_6/prodata/1/1aa/PCITUS COMB.pep:*
5: /cgn2_6/prodata/1/1aa/RE COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1617	100.0	1208	2	US-09-463-702A-2
2	1617	100.0	1208	2	US-09-699-135-2
3	108.5	6.7	762	2	US-09-252-991A-29423
4	107	6.6	1029	2	US-09-829-769-7
5	107	6.6	1029	2	US-10-033-301-22
6	102.5	6.3	279	2	US-09-252-991A-18933
7	102	6.3	817	2	US-09-252-991A-25887
8	99.5	6.2	512	2	US-09-252-991A-22339
9	94.5	5.8	208	2	US-09-127-928-2
10	94.5	5.8	1055	2	US-09-502-540-15453
11	92.5	5.7	515	2	US-09-252-991A-19259
12	92	5.7	416	2	US-09-808-387-38
13	91.5	5.7	2548	2	US-09-172-422-1
14	90	5.6	490	2	US-09-252-991A-18962
15	89.5	5.5	565	2	US-09-008-481A-6
16	89.5	5.5	565	2	US-09-195-666A-5
17	89.5	5.5	565	2	US-09-195-666A-49
18	89.5	5.5	565	2	US-09-309-592-6
19	89.5	5.5	565	2	US-09-635-705-5
20	89.5	5.5	565	2	US-09-635-705-49
21	89.5	5.5	565	2	US-09-634-858A-5
22	89.5	5.5	565	2	US-09-634-858A-49
23	89.5	5.5	565	2	US-08-869-827C-5
24	89.5	5.5	565	2	US-08-869-827C-49
25	89	5.5	1012	2	US-08-811-481-16
26	89	5.5	1012	2	US-09-876-527-16
27	88.5	5.5	600	2	US-09-252-991A-16741

28	88	5.4	274	2	US-09-252-991A-27466	Sequence 27466, A
29	88	5.4	3816	2	US-09-428-517-3	Sequence 3, App11
30	87.5	5.4	518	2	US-09-252-991A-27664	Sequence 27664, A
31	86.5	5.3	290	2	US-09-252-991A-27484	Sequence 27484, A
32	86.5	5.3	915	2	US-09-252-991A-23637	Sequence 23637, A
33	86	5.3	998	2	US-09-252-991A-24402	Sequence 24402, A
34	86	5.3	1735	2	US-09-902-540-14547	Sequence 14547, A
35	86	5.3	1747	2	US-09-902-540-14765	Sequence 14765, A
36	85	5.3	989	2	US-09-199-637A-273	Sequence 273, App
37	85	5.3	1190	2	US-09-902-540-12293	Sequence 12293, A
38	85	5.3	1673	2	US-09-418-710-70	Sequence 70, App1
39	85	5.3	1673	2	US-09-839-479-69	Sequence 69, App1
40	85	5.3	1674	2	US-09-418-710-1	Sequence 1, App11
41	85	5.3	1674	2	US-09-839-479-1	Sequence 276, App1
42	85	5.3	1681	2	US-10-037-417-20	Sequence 20, App1
43	85	5.3	1697	2	US-10-037-417-68	Sequence 68, App1
44	84.5	5.2	457	2	US-09-252-991A-28824	Sequence 28824, A
45	84.5	5.2	709	2	US-09-252-991A-21993	Sequence 21993, A

ALIGNMENTS

```
RESULT 1
US-09-463-702A-2
; Sequence 2, Application US/09463702A
; Patent No. 6335435
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO, AKIRO
; APPLICANT: KITAO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAL150
; CURRENT APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-463-702A-2
Query Match      100.0%; Score 1617; DB 2; Length 1208;
Best Local Similarity 100.0%; Pred. No. 2.4e-178;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QLTVALDMPREAIETLLCYELAPHHMLELATTYTHCRNLCPGGPAQLALAHRCPL 60
DB 907 QLTVALDMPREAIETLLCYELAPHHMLELATTYTHCRNLCPGGPAQLALAHRCPL 966
QY 61 AYLCAOQLPDPDQGGSSVVEPFMVKLVDSMGWELASVPRALCOLQWDHPRFGVGRGTGV 120
DB 967 AYLCAOQLPDPDQGGSSVVEPFMVKLVDSMGWELASVPRALCOLQWDHPRFGVGRGTGV 1026
QY 121 LVEFSLAFHLSFGDLTAEBKQICDPLXGVRQERQALRLRTFOAFHSVAPSPSCG 180
DB 1027 LVEFSLAFHLSFGDLTAEBKQICDPLXGVRQERQALRLRTFOAFHSVAPSPSCG 1086
QY 181 PCLEQDEERSRLKDLGRYFEEBEGQEPGMEDAQEPGQARIQDMEDVRCDIRQF 240
DB 1087 PCLEQDEERSRLKDLGRYFEEBEGQEPGMEDAQEPGQARIQDMEDVRCDIRQF 1146
QY 241 LSLRPEKSSRAVARIIFGIGSPCYPAOVYQODRRFMKLYLHLSFHALVGLATELLQV 300
DB 1147 LSLRPEKSSRAVARIIFGIGSPCYPAOVYQODRRFMKLYLHLSFHALVGLATELLQV 1206
```

QY 301 AR 302
Db 1207 AR 1208

RESULT 2

US-09-699-135-2
; Sequence 2, Application US/09699135
; Patent No. 6472513
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; APPLICANT: HIRAKI AND ASSOCIATES
; APPLICANT: SHIMAMOTO, AKIRO
; APPLICANT: KITAO, SAORI
; APPLICANT: FURUICHI, YASUHIRO
; TITLE OF INVENTION: HUMAN GENE RECO4 ENCODING HELICASE
; FILE REFERENCE: HIRAI150
; CURRENT APPLICATION NUMBER: US/09/699,135
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: US/09/463,702A
; PRIOR FILING DATE: 2000-01-24
; PRIOR APPLICATION NUMBER: PCT/JP98/03114
; PRIOR FILING DATE: 1998-07-10
; PRIOR APPLICATION NUMBER: JAPAN 9/200387
; PRIOR FILING DATE: 1997-07-25
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-699-135-2

Query Match 100.0%; Score 1617; DB 2; Length 1208;
Best Local Similarity 100.0%; Pred. No. 2,4e-178;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVALDMPREAIETLLCYELHPHMLLATTYTCRANCPCGPAQLALHRCPL 60
Db 907 QLTVALDMPREAIETLLCYELHPHMLLATTYTCRANCPCGPAQLALHRCPL 966
QY 61 AVCLAQQLPBDPGCGSSSVFEDMYKLVDSMGMEIASVRLCOLQMDHEPRTGVRRGTGV 120
Db 967 AVCLAQQLPBDPGCGSSSVFEDMYKLVDSMGMEIASVRLCOLQMDHEPRTGVRRGTGV 1026
QY 121 LVFSESLAFHRSFGDLTAEEKDQICDPLVGRVQAREQALRLRTFQAFHSVAFPSCG 180
Db 1027 LVFSESLAFHRSFGDLTAEEKDQICDPLVGRVQAREQALRLRTFQAFHSVAFPSCG 1086
QY 181 PCLEQODEERSTRKDLGRYFEEBEGEPGMEIDACGPEFGARLQDMEDQVRCIDTRQF 240
Db 1087 PCLEQODEERSTRKDLGRYFEEBEGEPGMEIDACGPEFGARLQDMEDQVRCIDTRQF 1146
QY 241 LSLRPEKFSRAVARIFHGIGSPCYPAQVYGODRRFMRKYLHSPALVGLATEBLLQV 300
Db 1147 LSLRPEKFSRAVARIFHGIGSPCYPAQVYGODRRFMRKYLHSPALVGLATEBLLQV 2206
QY 301 AR 302
Db 1207 AR 1208

RESULT 3

US-09-252-991A-29423
; Sequence 29423, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29423
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29423

Query Match 6.7%; Score 108.5; DB 2; Length 762;
Best Local Similarity 23.9%; Pred. No. 0.00311;
Matches 72; Conservative 21; Mismatches 79; Indels 129; Gaps 14;

QY 46 GPAQLALAH---RCPLAVCLAQQLPBDPGCGSSSVFEDMYKLVDSMGMEIASVRLALC 102
Db 470 GPSLGGPRRRHPRRRRPLA-----ETPGQRRSAEH-----APGTGSD----- 507
QY 103 QLOMDHEPRTGVRRGTGVLVFSESLAFHRSFGDLTAEEKDQICDPLVGRVQAREQA-- 160
Db 508 ---DHGPRDPGR-----DRPGLVQSRRRPRSCAGQGVHRRRLQADG 547
QY 161 -----LARLRFGAFHSVAFPSCGPCLQODEERSTRKDLGRYFEEBEGEP 210
Db 548 GRPRATGRPLRLR-----RPGRALPSRTER--RLAQVLGMRWRDPRRAP 594
QY 211 GGMEDA-----QGP-----EPGARLQDMEDQV 233
Db 595 GGRLARRDGHPRRRQAAGMFLRCFVPHQGGGQDARQOALQLRPLRLQRLDGRV 654
QY 234 RCDIRQLSLRPEKFSRAVARIFHG-----IGSPCY-----PAQVQODR 275
Db 655 QDQLPRRIP-RAURRRGRRAASAGEPRATDQPHRHHRVGDPHHFGGRPA---CQPR 710
QY 276 R 276
Db 711 R 711

RESULT 4

US-09-929-769-7
; Sequence 7, Application US/0929769
; Patent No. 6914130
; GENERAL INFORMATION:
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Polakis, Paul
; APPLICANT: Shou, Jianyong
; APPLICANT: Smith, Victoria
; APPLICANT: Soriano, Robert
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: P5007R1-US
; CURRENT APPLICATION NUMBER: US/09/929,769
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/089,653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/090,355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/104,257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/141,037
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: PCT/US99/12252

```

; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: PCT/US00/00376
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/20118
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/888,257
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-929-769-7
```

```

Query Match      6.6%; Score 107; DB 2; Length 1029;
Best Local Similarity 26.2%; Pred. No. 0.0072;
Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;
```

```

QY 29 LEIATTTTCGTCGCGAQAQALAHRCPLAVCIAQOL-----PEDPGQGS--SSV 79
DB 245 LQALMTL-----LSSPHGALVMSM-HRSHPLACPLRLQCYQKCVPDPTGFSLSFLKV 298
QY 80 BFDWVKLVDSMGWELASVRRALCOL--QMDHPRRTGVRRTGTVLVEFSLAH----- 130
DB 239 ILQMTLQWLDSPVEGGLPLAQLRMLASQASGRRLSDVNG-GILRLAEALAFRODLEVV 357
QY 131 -----LRSRGLTAEKXQIQIDFLYGRVQARERQALRLRTFQAFHVA--FPSC 179
DB 358 STVRVIAATLRS-GEQCSVEPDILISKVLQGLIEVNSPHLEILLTAFSATADASPFPAC 416
QY 180 GPCLSEQDEERSTRLLKDLGRYFEERE--GOEP--GMEDAQGEPEQARLQDWDQV 233
DB 417 KVVVV-----VSSLL--LQEBEPLAGKRGADGSLRAVRLGSPSSGILVWLEML 464
QY 234 RCDIROFTSLRBEK--FSSRAVARIFHGIGSPCYPAQVYQODRRFRKTY----- 281
DB 465 DEEV--VSSCDPLQLRLIFSR--KKGQAQVPS-----FRPYLLTLFTHQ 506
QY 282 -----LHLSFHALVGLATEE 296
DB 507 SSMPTLHOCIRVLLGKSRQ 526
```

```

RESULT 5
US-10-033-301-22
; Sequence 22, Application US/10033301
; Patent No. 6930172
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Guirey, Austin L.
```

```

; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C6
; CURRENT APPLICATION NUMBER: US/10/033,301
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-301-22
```

```

Query Match      6.6%; Score 107; DB 2; Length 1029;
Best Local Similarity 26.2%; Pred. No. 0.0072;
Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;
```

```

QY 29 LEIATTTTCGTCGCGAQAQALAHRCPLAVCIAQOL-----PEDPGQGS--SSV 79
DB 245 LQALMTL-----LSSPHGALVMSM-HRSHPLACPLRLQCYQKCVPDPTGFSLSFLKV 298
QY 80 BFDWVKLVDSMGWELASVRRALCOL--QMDHPRRTGVRRTGTVLVEFSLAH----- 130
```

```

Db      299  LKQLMWDSPGCVGPGPRAQLMFLASQASAGRLSDVRG--GLRLMALAFRODLEVS 357
QY      131  -----LRSPDDLTAEEKQICDPLYGVARERQALRLRTFOAFHSA--FPSC 179
Db      358  STVAVATATRS--GEQCSVEPDLISKVLQGLIEVRSPLIEELITLAFSATDAASPFPAC 416
QY      180  GPCLEQODEERSTRLKDLGRYFEEB--GQEP---GGMEDAGPEPGQARLQDMEQV 233
Db      417  KPVVV-----VSSIL--LQEEPLAGKPGADGSGLEAVRLGSSGLVDMLEML 464
QY      234  RCDIRQFLSLRPERK---FSSRAVARLFGIGSPCYPAQVYGODRRFRKY----- 281
Db      465  DPEV---VSSCPDLQLRLLFRR-----KKGQAQVPS-----FRPYLLTLPTHQ 506
QY      282  ----LHLSFHALVGLATEE 296
Db      507  SSWPTLHQICIRVLIGKSRQ 526

```

RESULT 6

```

US-09-252-991A-18933
; Sequence 18933, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18933
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18933

```

```

Query Match      6.3%; Score 102.5; DB 2; Length 279;
Best Local Similarity 25.8%; Pred. No. 0.0034;
Matches 64; Conservative 35; Mismatches 106; Indels 43; Gaps 11;

```

```

QY      59  PLAVCLAQQLPEDEGSSSVFDMVKLVDSGMELASV---RRALCOLQMDHEPRTGVR 115
Db      5  PLPV--EEQLLVAAHQGHRIEFD--GRADTLGGDDADRIDRQPEPELQADAQ----- 55
QY      116  RGTGVVVEPSLAFHLASPGDLTAEEKQICD--FLYGRVQAREQALRLRTFOAFHS 173
Db      56  ----LADVAF--BHVHPEBDHPQSRREQLHDROQRHQRECCRAQAFAEBEKKEEQGED 108
QY      174  VAFPSGCPGLEQODEERS--TRLKDL-----GRYEBEESGPGGMDAG 218
Db      109  DG--EIDPGVEQYDQDARADLQOQVGLDHALAARDPFEQAPGQAGAEITDVG 166
QY      219  P--EPQARLQDN--EDQVRCDIRQFLSLRPERKFSRAVARLFGIGSPCYPAQVYGQ 273
Db      167  QAVVDPRQARAHNLGEBHRVDDHRQRTERHPRQAEQGPVAVARLEMLDAENENTTAPQ 226
QY      274  DRFRWKY 281
Db      227  GTRQAKKH 234

```

RESULT 7

```

US-09-252-991A-25887
; Sequence 25887, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25887
; LENGTH: 817
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25887

```

```

Query Match      6.3%; Score 102; DB 2; Length 817;
Best Local Similarity 28.6%; Pred. No. 0.019;
Matches 67; Conservative 23; Mismatches 62; Indels 82; Gaps 16;

```

```

QY      49  QQLAHRCRPLAVCLAQQLPEDEGSSSVFDP-----MVKLVDSGMELASV----- 97
Db      299  EAQRRIHR--PLAVHLA--VPVDFGQAGFPRAESDRQFLAARLIDRM--QDAQVFDDRLA 352
QY      98  -----RRALCOLQMDHEPRTGVR-----GTGVL-----VESSELAFHRS- 133
Db      353  FRGVGQRAQRAFGQFLAAH--PRRGVQRAATVAEDGAGLVERQHVGEPPAFHRRDP 411
QY      134  -----PGDLTAEEKQICDPLYGVARERQALRLRTFOAFHSAVAFPSGC 180
Db      412  ELATFRTRRSWGD--AAAREQPA--GGRDQRHQQRHQRHQRQ-----VAAGEQG 460
QY      181  PCLE---QODEERSTRLKDLGRYFEEBEGQEPGKEDA---GQPEPGQARL 226
Db      461  ERLEGHHYQOEDQOQADQD--VQRHF-----VGGILLPGAFHQGHNPVQGRIL 506

```

RESULT 8

```

US-09-252-991A-22339
; Sequence 22339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22339
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22339

```

```

Query Match      6.2%; Score 99.5; DB 2; Length 512;
Best Local Similarity 24.1%; Pred. No. 0.019;
Matches 63; Conservative 23; Mismatches 80; Indels 89; Gaps 12;

```

```

QY      108  HEPRTGVRGTGVVVEPSLAFHLASPGDLTAEEKQICDPLYGVARERQALRLRT 167
Db      99  HEVLAGYQR--LAVE-----HADAPVVLGVE-----LGGQODRALQGLARRRQ 142
QY      168  FOAF-----HVAVPSGCPGLEQODE-----RSTRKDLIGRYF-----EESQ-- 208
Db      143  FAVVVDLHDATGEMALGNLQHQREQAQAPGHPPLVTLRAFLVEDIGERRAQLPATQVQVVD 202
QY      209  -----EPGMDAGPEPGQARLQDMEQVQCD- 236

```

```
Db 203 LVGAODRCRVVHHHQAALGLSBAIGMVVDPGGLADQCGVEIGQARVVLASDQFDVDA 262
Qy 237 -----IRQFLSLRPEERF-----SSRAVA-----RIFHIGISPCYQAQVYGODRRF-- 277
Db 263 EAGADPEVEGQLRGRRRORFVGEIDRHHVACHPRTTHAAGAAETVEGIDILLVA 322
Qy 278 WRXYLHLSFHAL---VGLATE 295
Db 323 AQAQVHARAHLMDPALLATE 343

RESULT 9
US-09-127-928-2
Sequence 2, Application US/09127928
Patent No. 64721272
GENERAL INFORMATION:
APPLICANT: Deng, Gang
APPLICANT: Lin, Jiling-Huey
APPLICANT: Morser, Michael J
TITLE OF INVENTION: DNA Encoding a No. 647212721 Human Inhibitor-of-Apoptosis
TITLE OF INVENTION: Protein
FILE REFERENCE: DNA Encoding HIAP3
CURRENT APPLICATION NUMBER: US/09/127,928
CURRENT FILING DATE: 1998-07-31
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 298
TYPE: PRT
ORGANISM: Homo sapiens
US-09-127-928-2

Query Match 5.8%; Score 94.5; DB 2; Length 298;
Best Local Similarity 20.8%; Pred. No. 0.032;
Matches 62; Conservative 32; Mismatches 113; Indels 91; Gaps 12;

Qy 25 PHHWLELATYTHRCINLCPGPAOLQALHRCPLAVCLAQQLPEDPGQSSSVFPMV 84
Db 16 PSHMA-----AGDSPT-----EKGRSL-----GSPVGLDTC 45
Qy 85 KLVDMSGWEIASVRALCOLQMDHEBRTGVRGTG-----LVESSELAFLRSPG 135
Db 46 RAMDHYDQGLIQLRPLTESEEBEGANTLSRGAAPGMSSELRLASFYDMLTAEVPP 105
Qy 136 DITAE-----EKQOI-CDPLXGRVQARER-----QALARL-----RT 167
Db 106 BLIAAAGFPHTHGQDKVRCFCYCGIQSWKRGDPTWTHAKMPPSCQFLNRSGKGFVHS 165
Qy 168 FQAFHSVAFPSGQCLQODEERSTRLKDLG-----RYFEEBEGQEPGMDAQP 219
Db 166 VERTHSQILGSWDPEEPDAPVADSVASGYPELPTRRREVQSSAQEPGVSPEAQ 225
Qy 220 -----EPQARLQDMDQVRCDIROFLSLRPEEKSSRAVARIF-----HGIGSPCY 267
Db 226 RAMWVLEPFGAR-----DVEAQRLRQBERCKVCLDAVALVIFVPCGHVLCABCAP 277

RESULT 10
US-09-902-540-15453
Sequence 15453, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
```

```
SEQ ID NO 15453
LENGTH: 1055
TYPE: PRT
ORGANISM: Myxococcus xanthus
US-09-902-540-15453

Query Match 5.8%; Score 94.5; DB 2; Length 1055;
Best Local Similarity 26.3%; Pred. No. 0.21;
Matches 82; Conservative 40; Mismatches 103; Indels 87; Gaps 20;

Qy 7 LDMPEBAIE---TLTLYELHPPHMLLATYTHRCINLCPG-----GPA 48
Db 344 LDNLEQAVOHLPAVILGHWLQALPR--ARFLATSR--ALGLPGRLLDLAPLSPVEEGES 399
Qy 49 QLOALHRCPLAVCL-AQQLPEDPG-----QSSSVFPMVKLV--SMGWEIASVRA 100
Db 400 RLEELVNSD--AVRLPQARREARGIFELTAGAPVADIVRODDGIALAIELAAAMA 456
Qy 101 LCOLQMDHEBRTGVRGTGVLVEFSELAFLRSPGDLTAERKQICDPLXGRVQARERQA 160
Db 457 LLSVQLRERLT--RR-----FELLRKGRPDGARQAATLRGALDMSWMLPEBERTA 506
Qy 161 LARLR-----RTQAFHSV-AFPGCGPCLQODEERSTRLKDLGRYFEEBEGQEPGME 214
Db 507 LARCSVFPGFTLEAAEAVILGLPDGPVAV--LDVLQSLRVSIL-RVLEAE--LPG--- 557
Qy 215 DAQBPBQCARLQDMDQVRCDIROFLSLRPEEKSSRAVARIFHGIGSPCYPAQVYQD 274
Db 558 -----GDSRLQGYE-----SIROYAARLAER-----GVGNAAALAEHRAD- 593
Qy 275 RRFMRXYLHLSF 286
Db 594 ---W--VLSLAY 600

RESULT 11
US-09-252-991A-19259
Sequence 19259, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19259
LENGTH: 515
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19259

Query Match 5.7%; Score 92.5; DB 2; Length 515;
Best Local Similarity 23.2%; Pred. No. 0.12;
Matches 69; Conservative 40; Mismatches 105; Indels 83; Gaps 15;

Qy 48 AQQLALA--HRCPLAVCLAQQLPEDPGQSSSVFPMVKLVDSMGWEIASVRR----- 99
Db 96 AQQLALVGGRRALQAQVGHVLD-----QLVEVLDIVYHALQVLAQAVGIG 144
Qy 100 ALCOLQMDHEBRTGVRGTGVLVEFSELAFLRSPGDLTAERKQICDPLXGRVQARERQ 159
Db 145 GLDHLQAEAGPRH--RRAQVGHANHLALHRQGVVDVLGH-----MVEBGAQASHRV 195
Qy 160 ALARLRRTFOAHSAVPSGCPCL-----EQDEERSTRLKD-----LGR 200
Db 196 ATAHLDASTQA--AAGDPGRGLQAALLOTTHQVGDQADEAQAORAEQOQLRRIGV 253
```

```
QY 201 YFEEB-EGOEFGMEDAAGPERG-----QARLODMEDOVRC 236
D 254 HLEBQAOLOHFRVEDA-GEADADVIALQHRHVALHHSBALVVEVQULDL- QVARE 311
QY 237 IROFLSRPEEKFSRAVARIFHG-IGSPCYAQVYGODRRFWRKYLTLSF-HALVG 291
D 312 AEAIVADDP-----RKVVRLEFAGRVAHQFVHQVHGRFRQLADLHLTAHEHVLG 362

RESULT 12
US-09-808-387-38
; Sequence 38, Application US/09808387
; Patent No. 6962985
; GENERAL INFORMATION:
; APPLICANT: Kaia Palm
; APPLICANT: Toni's Timmusk
; APPLICANT: Cemines Research
; TITLE OF INVENTION: MAMMALIAN NEURALIZED FAMILY OF
; FILE REFERENCE: CEMRES.001A
; CURRENT APPLICATION NUMBER: US/09/808,387
; CURRENT FILING DATE: 2001-03-14
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 416
; TYPE: PR
; ORGANISM: Homo sapien
US-09-808-387-38

Query Match
Best Local Similarity 18.6%; Score 92; DB 2; Length 416;
Matches 54; Conservative 41; Mismatches 90; Indels 106; Gaps 12;

QY 22 ELHPEHMLLEL-----ATTYHCLNCPGSPAOLOALAHRCPPY-----AVCLAQ 67
D 141 EMKHSYLDAIRGLDVEASSYSNEQOLCPYAAAGCRRGDACVYLHGEVCEICRLQV 200
QY 68 L-PEDPQSGSSVFPDMVKLVDSMGWELASVRLCOLQMDHEPRTGVRGTGVLEFSE 126
D 201 LHPEDP-----BORKAHEKICMLTFEHE-----ME 225
QY 127 LAFLHREPGDLTAEXDOIC-----DPLYGRVOAREO-----ALARLRRTFQ 169
D 226 KAPRQA-----SQDKVCSICMEVILEKASASERRGILSNCHTYCLSIQMRCAE 278
QY 170 AFHSVAPSGPC-----LEQDE--ERSTRLDLIG-----RYFEEBEGQEP 210
D 279 QFENPIIKSCPECHVISEFVIPSYYWVEDQKQNELLEAFQKGKAKCKYFEGQKGTCP 338
QY 211 GGMEDAGPEPGQARLODMEDOVQCDIRQFLSLAPEEKFSRAVARIFHGT 261
D 339 FGSKCLYRHAYPDGRLABPE-----KPRKOLSSQGTVRFPNSV 376

RESULT 13
US-09-172-422-1
; Sequence 1, Application US/09172422A
; Patent No. 6300485
; GENERAL INFORMATION:
; APPLICANT: Adams, Arwen B.
; APPLICANT: Chiu, Choi Yung
; APPLICANT: Duhl, David
; APPLICANT: Gortman, Susan W.
; APPLICANT: Leng, Song
; APPLICANT: Sheffield, Val
; APPLICANT: Welch, Juliet
; TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED
; TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
; FILE REFERENCE: 200130.442
; CURRENT APPLICATION NUMBER: US/09/172,422A
; CURRENT FILING DATE: 1998-10-14
```

```
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2548
; TYPE: PR
; ORGANISM: Homo sapien
US-09-172-422-1

Query Match
Best Local Similarity 21.5%; Score 91.5; DB 2; Length 2548;
Matches 55; Conservative 38; Mismatches 76; Indels 87; Gaps 12;

QY 47 PAQOLAHRCPPAVCLAQQLPEDPG-----QGSSVFPDMVKLVDSM 90
D 1798 PDELAAY-HPTPLS-----PELPSCKEFEKENKPSPKARKKSVXISVAL-DSM 1848
QY 91 GWEIASVRLALCOLQMDHEPRTGVRGTGVLEFSEL-----AFHLSRPGDLTAEE--KDQ 144
D 1849 HMQDSTVQ-----IIASVSDIKSMDFLLKRYNDLNDSDSKDT 1887
QY 145 ICDFLYGRVOAREKQALARTFQAFHSVAFPSGCGLEQODEERSTRLDLGRYFEE 204
D 1888 LVDPVF-----KKALKERQINFSEYSSAL-----AMDGKISRYKDYALFEQI 1932
QY 205 BEGGEFGMEDAAGPERGQARLODMEDOVQCDIRQFLSLAPEEKFSRAVARIFHGISP 264
D 1933 LEKTRLEQSDSGESP-----VRVWVNTFVFLDEYWN-----EFKT-----SD 1972
QY 265 CYPAQVYGODRRFWRK 280
D 1973 CTANKVPTKTKRK 1988
```

```
RESULT 14
US-09-252-991A-18962
; Sequence 18962, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18962
; LENGTH: 490
; TYPE: PR
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18962
```

```
Query Match
Best Local Similarity 25.1%; Score 90; DB 2; Length 490;
Matches 76; Conservative 35; Mismatches 108; Indels 84; Gaps 18;

QY 38 HCLNCPGSPAOLOALAHRC-----PPLAVCLAQQL-----PEDPQSGSSVFPDMVKL 86
D 24 HLFVECPA--RQLOQLHHRDIDPALLAERLAQALREGFHAFROPREAAAFGVPTLQA 81
QY 87 VD-SMG--WEIASVRLALCOLQMDHEPRTGVRGTGVLEFSELAFLHRSRPGDLTAEXD 143
D 82 EDWTLGGVGLATYARAV-----GHOPROPRRRG-GIRLAABALA-----GNAGMFEERD 133
QY 144 QICDFLYGRVOAREKQALARTFQAFHSVAFPSGCGPLEQO-----DEERST 192
D 134 VLATL-----ADRRQQL-----GDVQAIHQVLEASGTCGQGVGLGRGNHPQVDVPTLV 184
QY 193 RLK-----DLGRY-----FEEREGEPGME--DAGPFGQARLODMEDOV 233
```

Db	185	RARPLQLLLDRAQQLDILRRGNARDIVVEGSAAGMTEILDALALRAGE--RTGLYTEQL	243
Qy	234	RCRIREFSLYRREKSSRAVAARIINGISPCFPAQVVGQDQRFRKTKYLNLSFPAVTGLA	293
Db	244	A--LEQLLNRR-----RAVGRNRYLVGARREVVQALAGN-----QFLAAGGLA	283
Qy	294	TEE	296
Db	284	TDQ	286

RESULT 15
US-09-008-481A-6

```

Sequence 6 Application US/09008481A
Patent No. 6087151
GENERAL INFORMATION:
APPLICANT: ARIO, Takeshi
APPLICANT: TANAI, Madoka
APPLICANT: TORIGOE, Kakuji
APPLICANT: KURIMOTO, Masaaki
TITLE OF INVENTION: DNA CODING FOR MAMMALIAN L-
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Releasee #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/008,481A
FILING DATE: 16-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/598,369
FILING DATE: 08-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 95-42564
FILING DATE: 08-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,981
REFERENCE/DOCKET NUMBER: ARIO=1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-008-481A-6

```

Query Match	5.5%;	Score 89.5;	DB 2;	Length 565;
Best Local Similarity	26.0%;	Pred. No. 0.32;		
Matches 65; Conservative	23;	Mismatches	93;	Indels 69; Gaps 13;

```

OY      4  YQALIMPEEALIEILLCYL-----ELHPHHMLIELLA-----TYTHGRILNCPG---- 45
Db      318 VSGLDWTSBALAKASIVYLGPFLSPERQELLAKDLRGEMTLPTADLH--QSSPPGSLTG 376
OY      46  -GPAQLGALAHNC-----PLAVCLAQQLPDPQSSSVSEFMDVXKLVDSWGME 93
Db      377 QGVAKRFSL-FGCGQEDSVODAVMPSLALALAHAGIELALQALMEIGSLD-RKLDNSGQT 434
OY      94  LASVRALQQLQMDHEPRTGVRGRGTGLVFESIELAPLHRSFGDLTAEKQOICDPLTYGRV 153

```

Db	435	LHFLVA	-----RNGRDG	-----	-----VTMLHKGMDVNAARDGLSPLLLA-V	473
Qy	154	QANERDLARKRTQAFHVSVPSCGCLFQD	-----	-----	-----ERSTTLKLLGRYEEEGQEPG	212
Db	474	QGHRRICILTKR	-----	-----	-----AGACTSPDILKDKGTLCRLAAR	519
Qy	213	MEDAGPEPG	222			
Db	520	QAGADLQDPG	529			

Search completed: December 27, 2005, 21:55:17
Job time : 14 secs

—

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:53:22 ; Search time 44.6 Seconds
(without alignments)
2829.248 Million cell updates/sec

Title: US-09-889-325-4_COPY_907_1208
Perfect score: 1617
Sequence: 1 QLTVALDMPERAEITLLCY.....HLSPHALVGLATEELLQVAR 302

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBSCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBSCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBSCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10_PUBSCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US11_PUBSCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US12_PUBSCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1617	100.0	1208	3	US-09-889-325-4
2	1617	100.0	1208	5	US-10-723-860-4382
3	1617	100.0	1208	5	US-10-882-405-4
4	1617	100.0	1208	5	US-10-756-149-5768
5	1144.5	70.8	1216	6	US-11-084-955-2
6	427	26.4	1579	6	US-11-097-143-16743
7	169	10.5	874	4	US-10-437-963-121617
8	145.5	9.0	206	4	US-10-425-114-68934
9	145.5	9.0	215	4	US-10-425-115-338629
10	107	6.6	1029	4	US-10-476-924-22
11	107	6.6	1029	3	US-09-929-769-7
12	107	6.6	1029	3	US-09-866-034-22
13	107	6.6	1029	4	US-10-033-246-22
14	107	6.6	1029	4	US-10-033-301-22
15	107	6.6	1029	4	US-10-033-326-22
16	107	6.6	1029	4	US-10-033-245-22
17	107	6.6	1029	4	US-10-033-223-22
18	107	6.6	1029	4	US-10-033-167-22
19	107	6.6	1029	4	US-10-033-244-22
20	107	6.6	1029	4	US-10-033-435-22
21	107	6.6	1029	4	US-10-032-990-22
22	107	6.6	1029	4	US-10-032-996-22
23	107	6.6	1029	4	US-10-033-396-22
24	107	6.6	1029	4	US-10-439-249-22
25	107	6.6	1029	4	US-10-692-072-22
26	107	6.6	1029	6	US-11-021-330-7
27	107	6.6	1029	6	US-11-021-329-7

28	100	6.2	552	3	US-09-833-245-1335	Sequence 1335, Ap
29	97.5	6.0	1548	4	US-10-369-493-6347	Sequence 6347, Ap
30	96	5.9	1462	4	US-10-437-963-184121	Sequence 184121, Ap
31	94.5	5.8	298	4	US-10-235-026-2	Sequence 2, Appl
32	94.5	5.8	298	4	US-10-807-897-27	Sequence 27, Appl
33	94.5	5.8	298	4	US-10-839-882-18	Sequence 18, Appl
34	94	5.8	600	4	US-10-369-493-60878	Sequence 20878, A
35	92.5	5.7	779	4	US-10-343-663A-25	Sequence 418, Appl
36	92.5	5.7	779	5	US-10-370-715B-418	Sequence 418, Appl
37	92.5	5.7	779	5	US-10-852-335A-165	Sequence 165, Appl
38	92	5.7	416	3	US-09-808-387-38	Sequence 38, Appl
39	91.5	5.7	2548	3	US-09-851-682A-1	Sequence 1, Appl
40	91	5.6	465	3	US-09-764-864-1278	Sequence 1278, Ap
41	90.5	5.6	545	3	US-10-369-493-19756	Sequence 19756, A
42	90	5.6	943	4	US-10-437-963-168658	Sequence 168658, A
43	90	5.6	974	4	US-10-369-493-21047	Sequence 21047, A
44	89.5	5.5	670	5	US-10-631-467-863	Sequence 863, App
45	89.5	5.5	690	5	US-10-631-467-862	Sequence 862, App

ALIGNMENTS

RESULT 1
US-09-889-325-4
; Sequence 4, Application US/09889325
; Publication No. US20040224312A1
; GENERAL INFORMATION:
; APPLICANT: AGES Research Institute, Co., Ltd.
; TITLE OF INVENTION: Gene causative of Rothmund-Thomson syndrome
; TITLE OF INVENTION: and its gene product
; FILE REFERENCE: AI-003PCT
; CURRENT APPLICATION NUMBER: US/09/889,325
; PRIOR APPLICATION NUMBER: JP 1999-11218
; PRIOR FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-889-325-4

Query Match 100.0%; Score 1617; DB 3; Length 1208;
Best Local Similarity 100.0%; Pred. No. 4.4e-151;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	QLTVQALDMPERAEITLLCYELHAPHMWLELLATYTHCRINCPGPAQQLAHRCPEL	60
DB	907	QLTVQALDMPERAEITLLCYELHAPHMWLELLATYTHCRINCPGPAQQLAHRCPEL	966
QY	61	AVCLAAQQLPEDPGQSSSVFEDMVTLVDSMGWELASVRALCOLQDWHDEPRTVGRCTGV	120
DB	967	AVCLAAQQLPEDPGQSSSVFEDMVTLVDSMGWELASVRALCOLQDWHDEPRTVGRCTGV	1026
QY	121	LVESFELAFHLSPPDLTAEEKDOI CDPLVGVQARBOALRLRTPOAFSVAPSCG	180
DB	1027	LVESFELAFHLSPPDLTAEEKDOI CDPLVGVQARBOALRLRTPOAFSVAPSCG	1086
QY	181	PCLEQODDERSTRLLKDLGRYFESEEGEPGGMEDAOGPQCARLQDMEQVRCIDIRQF	240
DB	1087	PCLEQODDERSTRLLKDLGRYFESEEGEPGGMEDAOGPQCARLQDMEQVRCIDIRQF	1146
QY	241	LSLRPEBKSSSAVAVRIFFGIGSPCYPAQVYQODRRFRKXHLHLSFHALVGLATEELLQV	300
DB	1147	LSLRPEBKSSSAVAVRIFFGIGSPCYPAQVYQODRRFRKXHLHLSFHALVGLATEELLQV	1206
QY	301	AR 302	
DB	1207	AR 1208	

RESULT 2

US-10-723-860-4382
; Sequence 4382, Application US/10723860
; Publication No. US20040253606a1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882, 0193, NPUS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429, 739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4382
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-4382

Query Match 100.0%; Score 1617; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 4,4e-151;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 60
DB QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 966
QY 61 AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 120
DB AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
QY 121 LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 180
DB LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 1086
QY 181 PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 240
DB PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 1146
QY 1087 LSLRPEKFSRAVARIFHGIGSPCYPAQYGGDRRFRWKYLIHSFALVGLATEBELLQV 300
DB 1147 LSLRPEKFSRAVARIFHGIGSPCYPAQYGGDRRFRWKYLIHSFALVGLATEBELLQV 1206
QY 301 AR 302
DB 1207 AR 1208

RESULT 3

US-10-882-405-4
; Sequence 4, Application US/10882405
; Publication No. US20040259147a1
; GENERAL INFORMATION:
; APPLICANT: AGENE Research Institute, Co., Ltd.
; TITLE OF INVENTION: Gene causative of Rothmund-Thomson syndrome
; FILE REFERENCE: AI-003PCT
; CURRENT APPLICATION NUMBER: US/10/882, 405
; PRIOR FILING DATE: 2004-07-02
; PRIOR APPLICATION NUMBER: US/09/889, 325
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: JP 1999-11218
; PRIOR FILING DATE: 1999-01-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-882-405-4

Query Match 100.0%; Score 1617; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 4,4e-151;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 60
DB QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 966
QY 61 AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 120
DB AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
QY 967 AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
DB AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
QY 121 LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 180
DB LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 1086
QY 181 PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 240
DB PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 1146
QY 1087 LSLRPEKFSRAVARIFHGIGSPCYPAQYGGDRRFRWKYLIHSFALVGLATEBELLQV 300
DB 1147 LSLRPEKFSRAVARIFHGIGSPCYPAQYGGDRRFRWKYLIHSFALVGLATEBELLQV 1206
QY 301 AR 302
DB 1207 AR 1208

RESULT 4

US-10-756-149-5768
; Sequence 5768, Application US/10756149
; Publication No. US20050181375a1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND
; FILE REFERENCE: file
; CURRENT APPLICATION NUMBER: US/10/756, 149
; PRIOR FILING DATE: 2004-01-12
; NUMBER OF SEQ ID NOS: 5818
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5768
; LENGTH: 1208
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-756-149-5768

Query Match 100.0%; Score 1617; DB 5; Length 1208;
Best Local Similarity 100.0%; Pred. No. 4,4e-151;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 60
DB QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 966
QY 907 QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 966
DB QLTVALDMPBEALETLLCYELHPHMLBLATTTTHCRINCPCGPAQOLALAHRCPL 966
QY 61 AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 120
DB AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
QY 967 AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
DB AVCLAQQLPBDPGGSSSVFEDWVKLVDSMGWELASVRALCOLQMDHEPRTGVRGTGV 1026
QY 121 LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 180
DB LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 1086
QY 1027 LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 1086
DB LVERSELAFLHRSBGDLTAEEKDOI CDPLYGRVQAREQALRLRRTFOAFHSAVPSCG 1086
QY 181 PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 240
DB PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 1146
QY 1087 PCLEQODEERSTRKDLIGRYFEEBEGQEPGMEADAGPEPGQARLDQMEDQVRCIDIRQF 1146
DB 241 LSLRPEKFSRAVARIFHGIGSPCYPAQYGGDRRFRWKYLIHSFALVGLATEBELLQV 300

Db 1147 LSLRPEKFTSSRAVARLPHGIGSECPYQVYQGDRRFWKYLHLSFAVLGLATELLQV 1206

QY 301 AR 302

Db 1207 AR 1208

RESULT 5
US-11-084-955-2

```

: PUBLICATION NO.: US20050183193A1
: GENERAL INFORMATION:
: APPLICANT: National Institute of Radiological Sciences
: TITLE OF INVENTION: Targeted knockout mouse for human Rothmund-Thomson syndrome and
: TITLE OF INVENTION: preparation method thereof
: FILE REFERENCE: Y1L-0610
: CURRENT APPLICATION NUMBER: US/11/084,955
: CURRENT FILING DATE: 2005-03-21
: PRIOR APPLICATION NUMBER: JP 2003-185409
: PRIOR FILING DATE: 2003-06-27
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 2
: LENGTH: 1216
: TYPE: prt
: ORGANISM: Mus musculus
: US-11-084-955-2

```

```

1 RESULT 6
2 US-11-097-143-16743
3 / Sequence 16743, Application US/11097143
4 / Publication No. US20050208558A1
5 / GENERAL INFORMATION:
6 / APPLICANT: Venter, J. Craig
7 / APPLICANT: et al.
8 / TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
9 / TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
10 / TITLE OF INVENTION: DROSOPHILA GENES.
11 / FILE REFERENCE: CL000728
12 / CURRENT APPLICATION NUMBER: US/11/097,143
13 / CURRENT FILING DATE: 2005-04-04
14 / PRIOR APPLICATION NUMBER: 60/157,832
15 / PRIOR FILING DATE: 1999-10-05
16 / PRIOR APPLICATION NUMBER: 60/160,121
17 / PRIOR FILING DATE: 1999-10-19
18 / PRIOR APPLICATION NUMBER: 60/161,932
19 / PRIOR FILING DATE: 1999-10-28
20 / PRIOR APPLICATION NUMBER: 60/164,769

```

```

? PRIOR FILING DATE: 1999-11-12
? PRIOR APPLICATION NUMBER: 60/173,383
? PRIOR FILING DATE: 1999-12-28
? PRIOR APPLICATION NUMBER: 60/175,693
? PRIOR FILING DATE: 2000-01-12
? PRIOR APPLICATION NUMBER: 60/184,831
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: 60/191,637
? PRIOR FILING DATE: 2000-03-23
? NUMBER OF SEQ ID NOS: 4308
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 16743
? LENGTH: 1579
? TYPE: prt
? ORGANISM: DROSOPHILA
US-11-097-143--16743

```

Query Match	26.4%	Score 427	DB 6	Length 1579
Similarity	31.1%	Pred. No. 1.2e-32		
Best Local				
Matches 95	Conservative 62	Mismatches 12	Indels 26	Gaps 6

```

/ RESULT 7
/ US-10-437-963-121617
/ Sequence 121617, Application US/10437963
/ Publication NO. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204366
/ SEQ ID NO 121617
/ LENGTH: 874
/ TYPE: PR1
/ ORGANISM: Oryza sativa
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(874)
/ OTHER INFORMATION: unsure at all Xaa locations
/

```

```
FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24624C.1.pep
US-10-437-963-121617

Query Match      10.5%; Score 169; DB 4; Length 874;
Best Local Similarity 20.9%; Pred. No. 2.4e-07;
Matches 68; Conservative 58; Mismatches 118; Indels 82; Gaps 12;

QY 1 QLTVALDMPREAELETLCYELHPHMLLATTYTHGRNLPCGPAOLQALAHRCPL 60
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 598 ELTSRKEDIKKEEVLLITLTQIEIDQCYIRLLPQSVYCTI-----YFHKTS PQ 646

QY 61 AVC-----LAOQLPBDPGSGSSVEPDMVKLVDSKMWELASVRALCQIOWDHEPRTGRR 116
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 647 LLDKXDLIRSVLNRSEMKDGHYVFDIPIRINDLKITNEY-----FDHLK----- 693

QY 117 GTGLVLEFS-ELAFHLRSPG-----DLTAEEKDQICDPLYGRVQAEERQALRLR 165
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 694 -----LKFSGEISFELKDPAYCYIIMRPDPFNLSAN-----LTKWLSVEVSSKISKLD 743

QY 166 RTFQAFHSVAFPSG-----GPCLEQODEERSTRLDLGRYEEEGEPGME 214
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 744 AMF-ALANFAVKGCKRTGCGSGSQTHTPCIQKIME-----YFSKDGTSENDC- 790

QY 215 DAQGPBEGQARLDQMEDQVRCIDIRQFLSLRPEKSSRAVARIFHGSPCYPAQVYGD 274
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 791 -----RTQLQKSSPFLQDVKIFIQSNSPAKFTPRAVARIMHGISSPAFVSATWSKN 842

QY 275 RRFWRKYLHLSFHALVGLATEELLQV 300
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 843 -HFWGRYVEVDPLVMEAKAEIVKL 867

RESULT 8
US-10-425-114-68934
; Sequence 68934, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yinhua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53113)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68934
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73192E07_FIL.pep
US-10-425-114-68934

Query Match      9.0%; Score 145.5; DB 4; Length 206;
Best Local Similarity 25.0%; Pred. No. 7.6e-06;
Matches 50; Conservative 32; Mismatches 69; Indels 49; Gaps 9;

QY 123 EFS-ELAFHLRSPG-----DLTAEEKDQICDPLYGRVQAEERQALRLRTPQAF 171
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 29 QFSGEISFELKDPACIYVILKKPDDLNALAD-----ITRWLSVEVSSKISKLDAMPD-L 82

QY 172 HSAVAFPSG-----GPCLEQODEERSTRLDLGRYEEEGEPGMEDAQPE 220
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 83 ANFAVKGCKRTDGGSGSQTHTPCIQK-----IIQVFSKNYSTSDSG--QCTOPQ 129

QY 221 PGQARLDQMEDQVRCIDIRQFLSLRPEKSSRAVARIFHGSPCYPAQVYGD RRFWRK 280
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 RGSFFLQ-----ADIKVFLKNSPAKFTPRAVARIMHGISSPAFVSATWSKN-HFWGR 181
```

```
QY 281 YLHLSFHALVGLATEELLQV 300
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 YLEVDFFVWMEAKAEIVKL 201

RESULT 9
US-10-425-115-338629
; Sequence 338629, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53122)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 338629
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_71997C.1.pep
US-10-425-115-338629

Query Match      9.0%; Score 145.5; DB 4; Length 215;
Best Local Similarity 25.0%; Pred. No. 8e-06;
Matches 50; Conservative 32; Mismatches 69; Indels 49; Gaps 9;

QY 123 EFS-ELAFHLRSPG-----DLTAEEKDQICDPLYGRVQAEERQALRLRTPQAF 171
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 38 QFSGEISFELKDPACIYVILKKPDDLNALAD-----ITRWLSVEVSSKISKLDAMPD-L 91

QY 172 HSAVAFPSG-----GPCLEQODEERSTRLDLGRYEEEGEPGMEDAQPE 220
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 92 ANFAVKGCKRTDGGSGSQTHTPCIQK-----IIQVFSKNYSTSDSG--QCTOPQ 138

QY 221 PGQARLDQMEDQVRCIDIRQFLSLRPEKSSRAVARIFHGSPCYPAQVYGD RRFWRK 280
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 139 RGSFFLQ-----ADIKVFLKNSPAKFTPRAVARIMHGISSPAFVSATWSKN-HFWGR 190

QY 281 YLHLSFHALVGLATEELLQV 300
; : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 191 YLEVDFFVWMEAKAEIVKL 210

RESULT 10
US-10-476-924-22
; Sequence 22, Application US/10476924
; Publication No. US20040152093A1
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry; Ding, Li;
; APPLICANT: BAUGHN, Mariah R.; LAU, Preeti G.;
; APPLICANT: YUE, Huidun; HAPALIA, April J.A.;
; APPLICANT: LEE, Ernestine A.; ISON, Craig H.;
; APPLICANT: BECHA, Shanya D.; GURUBAUN, Rajagopal;
; APPLICANT: EMERLING, Brooke M.; GRIFFIN, Jennifer A.;
; APPLICANT: TANG, Y. Tom; LU, Dying Anna M.;
; APPLICANT: YAO, Monique G.; CHAWLA, Nandinder K.;
; APPLICANT: RAMKUMAR, Jayalakshmi; GANDHI, Ameena R.;
; APPLICANT: LEE, Soo Yeun; RICHARDSON, Thomas W.;
; APPLICANT: YANG, Junning; ELIOTT, Vicki S.;
; APPLICANT: LU, Yan; THANGAVELU, Kavitha;
; APPLICANT: HE, Ann; AZIMZAI, Yalda;
; APPLICANT: RAUMANN, Brigitte E.; SWARNAKAR, Anita;
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0960 USN
; CURRENT APPLICATION NUMBER: US/10/476,924
```

```

; CURRENT FILING DATE: 2003-11-04
; PRIOR APPLICATION NUMBER: PCT/US02/14276
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US 60/288,598
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/291,776
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/292,172
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/293,564
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1532441CD1
US-10-476-924-22

```

```

Query Match      6.6%; Score 107; DB 4; Length 751;
Best Local Similarity 27.0%; Pred. No. 0.29;
Matches 70; Conservative 23; Mismatches 106; Indels 60; Gaps 10;

```

```

QY 11 BEATITLCYELAHMHMLIATTT-----HCLNCPGPAQLOALAHCPPIAVCLA 65
DB 159 BEARERQAQCPOHGEALRFLCQPCSQLCRECLD-----PHLDHCLPIAEAVR 210
QY 66 QQLPDPGSSSVFEDWKLVDMSGWELASVR-----ALCOLMD-----HEPRTG 113
DB 211 ARPGLEG-----LLAGVDNNLVELBAARVEKEALATLRQARVGTQVEBAAG 261
QY 114 VARGTGVVEFSELAFLHLSRPGDLTAEBKQICDPLYGVRQ-ARERQALAR----- 163
DB 262 VLR--ALLAQKQEVGQLRAHVEAABEARERLAELEGEGVAAARAAAFARVULSLGREA 319
QY 164 -----LRTFQAFFHSVAFPSGCPCLCEQODEERSTLK---DLIGYFEEBEGQEG 211
DB 320 EILSLEGAIAQRIRLOCGCPMAFPAPCLLPLEIHPGLDKNCHILRLSFEQOPQKXG 379
QY 212 GMEGA--OGPEPQARLOD 228
DB 380 GKQAGTGGESQSRERD 398

```

```

RESULT 11
US-09-929-769-7
; Sequence 7, Application US/09929769
; Publication No. US20030055224A1
; GENERAL INFORMATION:
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Polakis, Paul
; APPLICANT: Shou, Jianyong
; APPLICANT: Smith, Victoria
; APPLICANT: Soriano, Robert
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wu, Thomas D.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITL OF INVENTION: TREATMENT OF TUMOR
; FILE REFERENCE: F5007R1-US
; CURRENT APPLICATION NUMBER: US/09/929,769
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: 60/089,653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/090,355
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/104,257
; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10

```

```

; PRIOR APPLICATION NUMBER: 60/141,037
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/20111
; PRIOR FILING DATE: 1999-09-01
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; PRIOR APPLICATION NUMBER: PCT/US00/00376
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: PCT/US00/04342
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/20118
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 09/888,257
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 8
; SEQ ID NO 7
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-929-769-7

```

```

Query Match      6.6%; Score 107; DB 3; Length 1029;
Best Local Similarity 26.2%; Pred. No. 0.44;
Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;

```

```

QY 29 LELATTTTHCRINCPGPAQLOALAHRCPIAVCLAQUL-----PEDPGQGS--SSV 79
DB 245 LQALATL-----LSSPHGALVMSH-HRSHPLACPLRLQLOYQRCVQDITGFSSLFKXV 298
QY 80 EFDVVKLVDSMGWELASVRALCOL--QMDHEPRTGVRGTGVVEFSELAH----- 130
DB 299 LLQMLQWIDSPGVGEPPLRAQLBMLASQASAGRRSLDVRG--GLRLABALAFRQDLEVS 357
QY 131 -----LNSPGDLTAEBKQICDPLYGVRQVAREQALRLKRTFQAFFHSVA--FPSC 179
DB 358 STVRVIAITLRS-GRQCSVEPDLISKVLQGLEVASPHLEBITLTFFSATDADASPFPAC 416
QY 180 GPCLEQODEERSTRLKDLIGRYFEEB---GOEP--GMEGAQPEPQARLODMEQV 233
DB 417 KPVVV-----VSSLL--LQEEBPLAGKFPADGSLFAVVLGSSGLLVMLENL 464
QY 234 RCDIRQFLSLRPEEK-----FSSRAVARIFHGISBPCYQVYGDORRFWRKY----- 281
DB 465 DEPV---VSSCPDLQLRLLFSSR-----KKGQAQVDS-----FRPYLLTLFTHQ 506
QY 282 -----LHISFHALVGLATEE 296
DB 507 SSMPTLHQCIRVLALKSRQ 526

```

```

RESULT 12
US-09-866-034-22
; Sequence 22, Application US/09866034
; Publication No. US20030170864A1
; GENERAL INFORMATION:

```

```
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P293ORIC1
CURRENT APPLICATION NUMBER: US/09/866,034
CURRENT FILING DATE: 2001-05-25
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 22
LENGTH: 1029
TYPE: PRT
ORGANISM: Homo sapiens
US-09-866-034-22
```

Query Match 6.6%; Score 107; DB 3; Length 1029;

Best Local Similarity 26.2%; Pred. No. 0.44;

Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;

```
QY 29 LELIATYYTHCRINCPGPAQLQALAHRCPLAVCLAQOL-----PEDPGQS--SSV 79
DB 245 LQALATL-----LSPHFGALVMSM-HRSHFLACPLRLQCYQRCVQDPTGFSSLFLKV 298
QY 80 EFDWVKLVDSKMGWELASVRALCOL--QMDHEPRTGVRGTGVIVSESLAFH----- 130
DB 299 LQWTLQWLDSPGVGGPLRAQLRLMLASQASAGRRLSDVRG--GLIRLAEALAFRDLEVVS 357
QY 131 -----LRSPDDLRAEEKDQICDFLYGRVQARERQALRLRRTQAFHSVA--FPSC 179
DB 358 STYRAVATATLRS--GEQCSVEPDLISKVLQGLIEVRSPLHEELLTAFFSATADAASPFPAC 416
QY 180 GPCLEQDDEERSTLXKDLGRYFEEB--GQEP--GGMADGSPREGQARLDQWBEQV 233
DB 417 KPVVV-----VSSIL--LQEEPLACGKCGADGSGLEAVRLGSSGLLDWLEML 464
QY 234 RCDIRQPLSLRPERK---FSSRAVARIFHGIGSPCYPAQVYGODRRFRKRY----- 281
DB 465 DPEV---VSSCPDLQRLPLFSRR-----KKGQAQVPS-----FRYLLTLFTHQ 506
QY 282 -----LHLSFHALVGLATBE 296
DB 507 SSWPTLHQICIRVLGKSRQ 526
```

RESULT 13

US-10-033-246-22

Sequence 22, Application US/10033246

Publication No. US20020098505A1

GENERAL INFORMATION:

APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gao, Wei-Qiang
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.

```

TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P293ORIC12
CURRENT APPLICATION NUMBER: US/10/033,246
CURRENT FILING DATE: 2001-12-28
PRIOR APPLICATION NUMBER: 60/095,325
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/112,851
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,145
PRIOR FILING DATE: 1998-12-16
PRIOR APPLICATION NUMBER: 60/113,511
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/115,558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115,733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119,341
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119,537
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/119,965
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: 60/162,506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170,262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/187,202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: PCT/US99/12252
PRIOR FILING DATE: 1999-06-02
PRIOR APPLICATION NUMBER: PCT/US99/28634
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US00/03565
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: PCT/US00/04414
PRIOR FILING DATE: 2000-02-22
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/08439
PRIOR FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: PCT/US00/14941
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
NUMBER OF SEQ ID NOS: 38
SEQ ID NO 22
LENGTH: 1029
TYPE: PRT
ORGANISM: Homo sapiens
US-10-033-246-22
```

Query Match 6.6%; Score 107; DB 4; Length 1029;

Best Local Similarity 26.2%; Pred. No. 0.44;

Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;

```
QY 29 LELIATYYTHCRINCPGPAQLQALAHRCPLAVCLAQOL-----PEDPGQS--SSV 79
DB 245 LQALATL-----LSPHFGALVMSM-HRSHFLACPLRLQCYQRCVQDPTGFSSLFLKV 298
QY 80 EFDWVKLVDSKMGWELASVRALCOL--QMDHEPRTGVRGTGVIVSESLAFH----- 130
DB 299 LQWTLQWLDSPGVGGPLRAQLRLMLASQASAGRRLSDVRG--GLIRLAEALAFRDLEVVS 357
QY 131 -----LRSPDDLRAEEKDQICDFLYGRVQARERQALRLRRTQAFHSVA--FPSC 179
DB 358 STYRAVATATLRS--GEQCSVEPDLISKVLQGLIEVRSPLHEELLTAFFSATADAASPFPAC 416
```

```
QY 180 GPCLEQDQERSTRKDLIGRYFEER---GOEP---GEMDAQGPPEGQARLODMEDQV 233
Db 417 KPVVV-----VSSLL---LQEEPLAGKPGADGSLFAVVLGSSGLVDWLEML 464
QY 234 RCDIRQPLSLRPEEK---FSSRAVARIFHGIGSPCYPAQVYGDRRFRMKY----- 281
Db 465 DPEV---VSSCPDQLRLLFSSR-----KKGQAQVPS-----FRPYLLTLFTHQ 506
QY 282 ----LHLSFHALVGLATEE 296
Db 507 SSWPTLHCIRVLGKSRQ 526
```

```
RESULT 14
US-10-033-301-22
; Sequence 22, Application US/10033301
; Publication No. US20020098506A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C6
; CURRENT APPLICATION NUMBER: US/10/033,301
; PRIOR FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
```

```
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-301-22
```

```
Query Match 6.6%; Score 107; DB 4; Length 1029;
Best Local Similarity 26.2%; Pred. No. 0.44;
Matches 84; Conservative 34; Mismatches 112; Indels 90; Gaps 18;
```

```
QY 29 LELATTYTHGRNCPCGPPAQDLAHRCPPLAVCLAQOL-----PEDGQGS--SSV 79
Db 245 LQALATL-----LSPFHGALVMSW-HRSFPLACPLRLQIQYQRCVPQDTGFSSSLFKV 298
QY 80 EFDWVKLVDSMGWELASVRAICOL--QWDHEPRTGVRGTVLVEFSELAFH----- 130
Db 299 LIQMLQWIDSPGVEGPIRLAQLMLASQASGRRLSDVRG-GLRLAALAFRODLEVVS 357
QY 131 -----LRSFGDLTAEEKQICDPLYGAVQAREQALRLRRTFOAFHSVA--FPSC 179
Db 358 STVRAVINTLRS-GRQCVSEPLDISKVLQGLIEVNSPHLEILTFFSATADAAPFPAC 416
QY 180 GPCLEQDQERSTRKDLIGRYFEER---GOEP---GEMDAQGPPEGQARLODMEDQV 233
Db 417 KPVVV-----VSSLL---LQEEPLAGKPGADGSLFAVVLGSSGLVDWLEML 464
QY 234 RCDIRQPLSLRPEEK---FSSRAVARIFHGIGSPCYPAQVYGDRRFRMKY----- 281
Db 465 DPEV---VSSCPDQLRLLFSSR-----KKGQAQVPS-----FRPYLLTLFTHQ 506
QY 282 ----LHLSFHALVGLATEE 296
Db 507 SSWPTLHCIRVLGKSRQ 526
```

```
RESULT 15
US-10-033-326-22
; Sequence 22, Application US/10033326
; Publication No. US20020098507A1
; GENERAL INFORMATION:
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2930R1C8
; CURRENT APPLICATION NUMBER: US/10/033,326
; PRIOR FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/095,325
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/112,851
; PRIOR FILING DATE: 1998-12-16
```

```
; PRIOR APPLICATION NUMBER: 60/113,145
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: 60/113,511
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/115,558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115,733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119,341
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: 60/119,537
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/119,965
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: 60/162,506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 60/170,262
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 60/187,202
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: PCT/US99/28634
; PRIOR FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US00/03565
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 38
; SEQ ID NO 22
; LENGTH: 1029
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-033-326-22
```

Query Match 6.6%; Score 107; DB 4; Length 1029;

Best Local Similarity 26.2%; Pred. No. 0.44; Mismatches 112; Indels 90; Gaps 18;

```
29 LELIATYTHCRNLCPGGPAQLALHRCPLAVLAQOL-----PEDPGGS--SSV 79
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
245 LQALATL-----LSPHGGALVMSM-HRSHFLACPLRLQCOYQRCVQDTPGSSLPKLV 298
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
80 EFDNVKLVDSKMGWELASVRRALCOI--QWDHEPRTGVRRGTGLVEFSELAFH----- 130
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
299 LQWMLQWLDSPGVGGPLFAQLRLMLASQASAGRRLSDVRG-GLRLLEALAFRODLEIVS 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
131 -----LRSPGDLTAEEKDOICDFLYGRVQAREQALRLRTRTQAFHSVA--PPSC 179
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
358 STYRAVAATLRS-GEQGSVEPDLISKVLQGLIEVRSPLHEELLTAFFSATADAASPPFAC 416
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
180 GPCLLEQDDEKSTRLKDLGRYFEER---GQEP---GMEDAQCPBPGQARLDQWEDQV 233
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
417 KPVVV-----VSSIL---LQEBEPLAGKFGADGSLBAVRLGSSGLVDWLEML 464
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
234 RCDIRQFLSLRPEEK---FSSRAVARIFFHGIGSPCYPAQVYGQDRRFMRKY----- 281
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
465 DPEV---VSSCPDQLRLRLFRR-----KKGQAQVPS-----FRPYLLTLFTHQ 506
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

OY 282 -----LHLSFHAIVGLATEE 296
DB 507 SSMPTLHQICIRVLGKSREQ 526

Search completed: December 27, 2005, 22:12:35
Job time : 45.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:54:12 ; Search time 3 Seconds
(without alignments)
718.036 Million cell updates/sec

Title: US-09-889-325-4_COPY_907_1208
Perfect score: 1617
Sequence: 1 QLVQALDMPBAITLLCY.....HLSPHATVGLATEELLQVAR 302

Scoring table: BLOSSUM62
Gapop 10.0 , Gapext 0.5

Searched: 54001 seqs, 7132810 residues

Total number of hits satisfying chosen parameters: 54001

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:*
1: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubppa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.5	5.5	3674	7 US-11-000-463-454	Sequence 454, App
2	88	5.4	3375	7 US-11-044-111-23	Sequence 23, App1
3	80.5	5.0	599	7 US-11-109-157A-3	Sequence 1, App1
4	80.5	5.0	1686	7 US-11-109-157A-1	Sequence 8056, Ap
5	79.5	4.9	717	6 US-10-467-657-8056	Sequence 940, App
6	78	4.8	615	6 US-10-995-561-940	Sequence 942, App
7	76	4.7	562	7 US-11-065-943-24	Sequence 941, App
8	75.5	4.7	662	6 US-10-995-561-943	Sequence 942, App
9	75.5	4.7	702	6 US-10-995-561-942	Sequence 941, App
10	75.5	4.7	754	6 US-10-995-561-941	Sequence 897, App
11	75	4.6	366	7 US-11-000-463-897	Sequence 425, App
12	75	4.6	400	7 US-11-000-463-425	Sequence 136, App
13	74	4.6	278	6 US-10-131-826A-136	Sequence 1658, Ap
14	74	4.6	777	6 US-10-821-234-1658	Sequence 773, App
15	74	4.6	3803	6 US-10-995-561-773	Sequence 771, App
16	74	4.6	3360	6 US-10-995-561-771	Sequence 774, App
17	74	4.6	5335	6 US-10-995-561-777	Sequence 779, App
18	74	4.6	5406	6 US-10-995-561-774	Sequence 775, App
19	74	4.6	5415	6 US-10-995-561-775	Sequence 776, App
20	74	4.6	5464	6 US-10-995-561-776	Sequence 65, App1
21	74	4.6	5935	6 US-10-995-561-775	Sequence 128, App1
22	73.5	4.5	545	7 US-11-065-943-65	Sequence 32, App1
23	73.5	4.5	562	7 US-11-065-943-28	Sequence 12, App1
24	73.5	4.5	562	7 US-11-065-943-32	
25	73.5	4.5	677	6 US-10-982-545-12	

26	73	4.5	798	7 US-11-107-028-3	Sequence 3, App1
27	73	4.5	1068	6 US-10-467-657-2904	Sequence 2904, Ap
28	72	4.5	545	7 US-11-065-943-67	Sequence 67, App1
29	72	4.5	562	7 US-11-065-943-34	Sequence 34, App1
30	71.5	4.4	596	7 US-11-082-389-420	Sequence 420, App
31	71	4.4	286	7 US-11-063-343-22	Sequence 22, App1
32	71	4.4	554	6 US-10-850-816-2	Sequence 2, App1
33	71	4.4	562	7 US-11-065-943-26	Sequence 26, App1
34	71	4.4	562	7 US-11-065-943-36	Sequence 36, App1
35	71	4.4	616	6 US-10-982-545-5	Sequence 5, App1
36	71	4.4	832	7 US-11-065-943-100	Sequence 100, App
37	71	4.4	832	7 US-11-007-797A-11	Sequence 11, App1
38	71	4.4	832	7 US-11-007-642B-11	Sequence 11, App1
39	70.5	4.4	562	7 US-11-065-943-22	Sequence 22, App1
40	70.5	4.4	1388	6 US-10-821-234-1143	Sequence 1143, Ap
41	70	4.3	545	7 US-11-065-943-97	Sequence 97, App1
42	70	4.3	691	6 US-10-131-826A-16	Sequence 16, App1
43	70	4.3	703	6 US-10-821-234-963	Sequence 963, App
44	70	4.3	1798	6 US-10-995-561-1033	Sequence 1033, Ap
45	70	4.3	1798	6 US-10-995-561-1034	Sequence 1034, Ap

ALIGNMENTS

RESULT 1
US-11-000-463-454
; Sequence 454, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radjole T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; PRIOR FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 454
; LENGTH: 3674
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-000-463-454

Query Match 5.5%; Score 88.5; DB 7; Length 3674;
Best Local Similarity 20.6%; Pred. No. 4.1;
Matches 68; Conservative 40; Mismatches 83; Indels 139; Gaps 16;
QY 6 ALDMPBAITETLLCYELHPHMLELATTYTRCRINCPGPAQLALHRCPLAVCIA 65

Db 823 ALSPGSLR-----NPGWSE-----ASCHPGSGAMKMA----- 853
Qy 66 QQLPBDPQSSSVFPMVKLV-----SMGWE-----LASVRA----- 100
Db 854 --LPAEP-----DDFDPNTLITQODHLSDYESTLRALAQRLRARBAMALFGPCSSCG 906
Qy 101 -----LQOLQMDHEPRTGVRGTGVLFVESELAFLHRS--P 134
Db 907 ELQJLMEKQVTLQVRPOADTLEWQJLKENF--LTLAVGKMAEVSASAEQLRQRP 965
Qy 135 GDLTAEEKDQ-ICDFLYGRVQARERQALARTFOAFHSAFSP-----CGPCTLEQODEE 189
Db 966 GNSFOIQROQEBELSGRWQLALNRE-----KANQLASHSVGCSFLQCCP----- 1011
Qy 190 RSTLKLGLGYFEEBEGQEPGEMDAG-----PEPGQ 223
Db 1012 TQVQLRDTL-----LQLEMLQPGSSSEDTLHALQLAQKTLVLERVYFLQSVVVKVEEPGY 1067
Qy 224 ARLODMEDQVRCDIROPLSLRPEKFSRA 253
Db 1068 AESQPLQGV--ETLQGLLKQVQEQVQRA 1095

RESULT 2

US-11-044-111-23
; Sequence 23, Application US/11044111
; Publication No. US20050272362A1
; GENERAL INFORMATION:
; APPLICANT: Chiang, Wen
; APPLICANT: Strasburg, Gale
; APPLICANT: Linz, John
; TITLE OF INVENTION: Genetic Test for PSE-Susceptible Turkeys
; FILE REFERENCE: MSU-09308
; CURRENT APPLICATION NUMBER: US/11/044,111
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 23
; LENGTH: 3375
; TYPE: PRT
; ORGANISM: Meleagris gallinavo
US-11-044-111-23

Query Match 5.4%; Score 88; DB 7; Length 3375;
Best Local Similarity 21.9%; Pred. No. 4.1; Indels 150; Gaps 17;
Matches 87; Conservative 35; Mismatches 126; Indels 150; Gaps 17;

Qy 4 VOALDMEE-----AIEPLLCTELPHPHMELATTYHCHLNCGGPAQOLAHRCRP 59
Db 230 MMALHIEBENRCHMDIMELMERQDLRFHM-----HTLKLKVCALGNTRVAAHALCSHYDP 285
Qy 60 LAVCLAQQLPBDP-----QGSSEVFPM----- 83
Db 286 SQLFAIRSPFELPQRLRAGYVDLLAVHLBQGVBARASMSSTEFIVPMDEASKRISLPAG 345
Qy 84 -----VKLVDSMGWELASV-----RRLCQLQMD 107
Db 346 GGGGVKVPGGPVGSLACLRPRHAEBCFVRPPDGRALLPSILRALGRARIMLR-- 403
Qy 108 HEPTG-----VRSGTVLVEFS-----ELAFHLSPPGDLTAEBKQDLDPLGYH-- 153
Db 404 -EAVAGGPHARDVGGVGEVQLVPLVKLVASALAVGLARPEYKVLMTIEPRVGGGE 462
Qy 154 QARERQALARTFOAFHSAFSPCGPCTLEQODEERSTRKLKDLGRYFEEBEGQEPGGM 213
Db 463 EEEEEERRRRRKAVEA-----GBEEBEVDEEBEV-----DEEBEB----- 499
Qy 214 EDAGGPPRG--QALQDMEDQVRCDIROF-----LSLPER--KFSRAVARIFHGIGSP 264
Db 500 EEBEGPEBGLQMKLPSSVKLQMCNLQFCDQELQHRVEAIVAFSEHVERLQR----- 554
Qy 265 CYPQAVYGODRRFMKRYLHLSFHALVGLATEELLQVAR 302

Db 555 -----DQRRRYGR-----LMGAVTMSAETAR 576

RESULT 3

US-11-109-157A-3
; Sequence 3, Application US/1109157A
; Publication No. US20050277175A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: TRUNCATED ADAMTS MOLECULES
; FILE REFERENCE: 01997.030500.
; CURRENT APPLICATION NUMBER: US/11/109,157A
; CURRENT FILING DATE: 2005-04-18
; PRIOR APPLICATION NUMBER: 60/562,685
; PRIOR FILING DATE: 2004-04-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 599
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-11-109-157A-3

Query Match 5.0%; Score 80.5; DB 7; Length 599;
Best Local Similarity 23.2%; Pred. No. 2.5; Indels 155; Gaps 20;
Matches 82; Conservative 36; Mismatches 81; Indels 155; Gaps 20;

Qy 44 PGGAQOLALAHRCPPPLAVCIAQQLPBDP-----QGSSEVFPMVKLVDSMG----- 91
Db 2 PGSPSP-RSPAPLRLPILLILLCALAPGAPGAPRATEGRAALDIYHPRVDAGGSFLSY 60
Qy 92 --WELA-----SVRR--ALQOLQMDHEPRTGVRGTGVLFVESELAF-----HURSP 134
Db 61 ELMPRLARKRDVSVRRDAPAFYELQY-----RG-----RELPFLMTANQHLAP 104
Qy 135 G-----DLTAEBKD-----QICDPLXGRVQ----- 154
Db 105 GFVSETRRGGLGRAHTRANTPACHLGEVDDPELBGLAIAISACDGLKGVQLSNEYF 164
Qy 155 -----AR-----ERQALARTFOAFHSAFSPCG-----PCV-----EQ 185
Db 165 IBPLDSAPARGHAGHPVVYKQAPERLA---QKGDSSASSTCGVQYVPELESRRERMEQ 221
Qy 186 QDEERSTRKLKDLGRYFEEBEGQEPGEMDAGPE--PGQARLODMEDQVRCDIROPLSLR 244
Db 222 RQWRPRRLRLRHQRSVKSKWVETLVADAKWVEYHGQFQVESY-----VLTIMMM- 273
Qy 245 PEKFSRAVARIFH--GISPCYPQAVYGODRRFMKRYLHLSFHALVGLATEE 296
Db 274 -----VAGLPHDPSIGNP-----IHTTVRLVLEDEE 301

RESULT 4

US-11-109-157A-1
; Sequence 1, Application US/1109157A
; Publication No. US20050277175A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: TRUNCATED ADAMTS MOLECULES
; FILE REFERENCE: 01997.030500.
; CURRENT APPLICATION NUMBER: US/11/109,157A
; PRIOR APPLICATION NUMBER: 60/562,685
; PRIOR FILING DATE: 2004-04-15
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 1686
; TYPE: PRT
; ORGANISM: homo sapiens
US-11-109-157A-1

Query Match 5.0%; Score 80.5; DB 7; Length 1686;

[illegible]

```

RESULT 5
US-10-467-657-8056
; Sequence 8056, Application US/10467657
; Publication No. US20050260581A1
GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/467,657
CURRENT FILING DATE: 2003-08-11
PRIOR APPLICATION NUMBER: GB-0103424.8
PRIOR FILING DATE: 2001-02-12
NUMBER OF SEQ ID NOS: 9218
SOFTWARE: SeqMan99, version 1.04
SEQ ID NO 8056
LENGTH: 717
TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8056

```

	Query Match	4.9%	Score 79.5;	DB 6;	Length 717;	
	Best Local Similarity	21.5%	Pred. No. 4;			
	Matches	39;	Conservative	28;	Mismatches	69;
					Indels	45;
					Gaps	5
<hr/>						
Qy	47	PAQLGALAHRCPEPLAVCLAAQQLPEDPGQSSSVFEDWVKLVDSMGWEILASVRALCOLQM	106			
		:: :				
Db	447	FVQEYAPADDDAKTILGGATVLPEPNGNGVVWVIDITVL-----RAQKEAM	495			
		:: :				
Qy	107	DHEPTGVRGRGVLFVEFSLEAFHLRSPGDLLAEKKDQCFLYRGVAAREKOLAR---	163			
		:: :				
Db	496	GEVAK-----RLHETIRNPITPQLSAERIAMLGGGLDDODAOILTRSTD	541			
		:: :				
Qy	164	-----LRRTFOAHSVAPFSCPCCEQQODEERSTRKLKLLGY-----FEEEGOE	209			
		:: :				
Db	542	TIIKQVALKEMVEAFRNVA---RAPSIKLEQNODINALIGDVALLYEAGPCRFAELAGE	598			
		:: :				
Qy	210	P 210				
Db	599	P 599				

```

RESULT 6
US-10-995-561-940
: Sequence 940, Application US/10995561
: Publication No. US20050272054A1
:
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele et al.
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
: TITLE OF INVENTION: DETECTION AND USES THEREOF
: FILE REFERENCE: CL001559
: CURRENT APPLICATION NUMBER: US/10/995,561
: CURRENT FILING DATE: 2004-11-24
:
: NUMBER OF SEQ ID NOS: 85702
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 940
: LENGTH: 615
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
US-10-995-561-940

```

Query March	4.8%;	Score 78;	DB 6;	Length 615;
Best Local Similarity	28.1%;	Pred. No. 4.6;		
Matches	38;	Conservative 15;	Mismatches 58;	Indels 24; Gaps 6
QY	22	ELIPIHMLIYATYTHCRLNCPGPAOL---	QALAHRCPLVAVCLAQOLPEDG----	73
Db	495	EVGVSDMLIFPASTLGSFLSILLVGLVGLYGLINRAARHICPPLPPICASSAIEFPFGKETW		554
QY	74	QGSSSYVEPDMVKLVDSMGWELASVYRAL--CQLQMDHERTCGVKRGTVLVVFSFLAHLR		1324
Db	555	QMINPVPDFQ-----EEASLQELVALVEKSMWDKGERTEPLEKT---	ELIPGARELA	6004
QY	133	SPGDLTAEEKDQICD	147	
Db	601	LPTLELSLEDDR-CD	614	

```

RESULT 7
US-11-065-943-24
: Sequence 24, Application US/11065943
: Publication No. US20050250131A1
: GENERAL INFORMATION:
: APPLICANT: JESTIN, JEAN-LUC
: APPLICANT: VICHIER-GUERRE, SOPHIE
: APPLICANT: FERRIS, STEPHANE
: TITLE OF INVENTION: METHODS FOR OBTAINING THERMOSTABLE ENZYMES, DNA POLYMERASE I
: TITLE OF INVENTION: VARIANTS FROM THERMUS AQUATICUS HAVING NEW CATALYTIC ACTIVITIES
: TITLE OF INVENTION: METHODS FOR OBTAINING THE SAME, AND APPLICATIONS OF THE SAME
: FILE REFERENCE: 266426USOXCIP
: CURRENT APPLICATION NUMBER: US/11/065,943
: CURRENT FILING DATE: 2005-02-25
: PRIOR APPLICATION NUMBER: US 10/787,219
: PRIOR FILING DATE: 2004-02-27
: NUMBER OF SEQ ID NOS: 106
: SOFTWARE: PatentIn version 3.3
: SEQ ID NO 24
: LENGTH: 562
: TYPE: PRT
: ORGANISM: Thermus aquaticus
US-11-065-943-24

```

```

Query Match 4.7% ; Score 76 ; DB 7 ; Length 562 ;
Best Local Similarity 21.9% ; Pred. NO. 6.3 ;
Matches 70 ; Conservative 39 ; Mismatches 136 ; Indels 74 ; Gaps 14 ;

QY 4 VOALDMPBEAETLLCYVEL-----HHHMLLELATYTHORLN----- 42
Db 242 LEALREAPRYEKLQYRELNKSTOTQLDPLNPR-----TSLHRTFNQATOT 294
QY 43 --CPGPAQOLALHRCPPPLAVCAQQLPEDPGQSSSVFERDQKLVDSMGWELASVRR 100
Db 235 GRSSSQPNLNINIPR-TPIGQIRRTETIARSGRLVLDVNOQL-----RV 341

```

```
Qy 101 LCOLQMDHEPRTGVRGTGLVVEFSELAFHL-RSPGDLTAEEKQICPF--LYGRVQARE 157
Db 342 LAHLISGSDNLIRVQEGSDIHETSMFVGPREAVDPLMRAAKTINFGVLYGMSARL 401
Qy 158 RQALARLRRTFOAF--HSVAFPSGCGPCLBQODEERSTR--LKDLLG--RYFEEEGQEP 210
Db 402 SQEALIPYEEAQAFIERFYQSGFFKVRAMIEKLEBGRRGVETLFGRRRYLPDLAQVK 461
Qy 211 GGMEDAQ-----PERGA-----RLQDMEDQVRCDIROFLSLR-PEEKSS 251
Db 462 NVREAEERARFNMVQGTAAADMLKLAIVKLFPRLEEMGARMLQVHDELVLAEAPKE--GA 519
Qy 252 RAVARIFHGIGSPCYAV 270
Db 520 EAVARLAEVMEGVYPLAV 538
```

```
RESULT 8
US-10-995-561-943
; Sequence 943, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 943
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-943
```

```
Query Match 4.7%; Score 75.5; DB 6; Length 662;
Best Local Similarity 27.1%; Pred. No. 8.7;
Matches 36; Conservative 16; Mismatches 58; Indels 23; Gaps 5;

Qy 21 LEIAPHHMLLEIATTTTHCRINCPGPAQL---QALAHRCPLAVCLAQQLPEDPG---- 73
Db 539 IEVQVSDMLIFPAGLSFSLILVGLVGLGNRAARHLCPPLTPPCASSAIEFFGKET 598
Qy 74 -QSSSVFEDNVKLVDMSGWELASVRRL-CQLQMDHEPRTGVRGTGLVVEFSELAFHL 131
Db 599 WQWINPVDFQ-----EASIQEALVVMESWDKGERTEPLEKT-----ELPEGABEL 644
Qy 132 RSPGDLTAEEKQ 144
Db 645 ALDTLESLDGDGR 657
```

```
RESULT 9
US-10-995-561-942
; Sequence 942, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 942
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-942
```

```
Query Match 4.7%; Score 75.5; DB 6; Length 702;
Best Local Similarity 27.1%; Pred. No. 9.3;
Matches 36; Conservative 16; Mismatches 58; Indels 23; Gaps 5;

Qy 21 LEIAPHHMLLEIATTTTHCRINCPGPAQL---QALAHRCPLAVCLAQQLPEDPG---- 73
Db 579 IEVQVSDMLIFPAGLSFSLILVGLVGLGNRAARHLCPPLTPPCASSAIEFFGKET 638
Qy 74 -QSSSVFEDNVKLVDMSGWELASVRRL-CQLQMDHEPRTGVRGTGLVVEFSELAFHL 131
Db 639 WQWINPVDFQ-----EASIQEALVVMESWDKGERTEPLEKT-----ELPEGABEL 684
Qy 132 RSPGDLTAEEKQ 144
Db 685 ALDTLESLDGDGR 697
```

```
RESULT 10
US-10-995-561-941
; Sequence 941, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 941
; LENGTH: 754
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: VARIANT
; LOCATION: (1)...(754)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-995-561-941
```

```
Query Match 4.7%; Score 75.5; DB 6; Length 754;
Best Local Similarity 27.1%; Pred. No. 10;
Matches 36; Conservative 16; Mismatches 58; Indels 23; Gaps 5;

Qy 21 LEIAPHHMLLEIATTTTHCRINCPGPAQL---QALAHRCPLAVCLAQQLPEDPG---- 73
Db 579 IEVQVSDMLIFPAGLSFSLILVGLVGLGNRAARHLCPPLTPPCASSAIEFFGKET 638
Qy 74 -QSSSVFEDNVKLVDMSGWELASVRRL-CQLQMDHEPRTGVRGTGLVVEFSELAFHL 131
Db 639 WQWINPVDFQ-----EASIQEALVVMESWDKGERTEPLEKT-----ELPEGABEL 684
Qy 132 RSPGDLTAEEKQ 144
Db 685 ALDTLESLDGDGR 697
```

```
RESULT 11
US-11-000-463-897
; Sequence 897, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhou, ping
US-11-000-463-897
```

```
APPLICANT: Cao, Yi-Cheng
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 785CIP4CN
CURRENT APPLICATION NUMBER: US/11/000,463
CURRENT FILING DATE: 2004-11-29
PRIOR APPLICATION NUMBER: 10/291,265
PRIOR FILING DATE: 2002-11-08
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 09/922,279
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/617,746
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/631,451
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 09/633,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 897
LENGTH: 366
TYPE: PRT
ORGANISM: Homo sapiens
US-11-000-463-897
```

```
Query Match 4.6%; Score 75; DB 7; Length 366;
Best Local Similarity 20.3%; Pred. No. 4.7;
Matches 46; Conservative 35; Mismatches 86; Indels 60; Gaps 8;
```

```
QY 54 AHCPCPLAVCLAQQLPEPDCGSSSVFEDMKLVDSMGHELASVRALCOLQMDHEPRFG 113
DB 82 APRLPQDPVGMKRQQLQEBLEEVKARLOPYMAEAEHVLGMNLGLRQQLKPYTMDLMEQVA 141
QY 114 VARGTGVLPSEFLAFHLRSPGDLTAEE---KDQICDFLYGRVOARERQALRLRTFQ 169
DB 142 LR-----VOELQELRVGSDTKAQLGVDAMALLQG-LQSRVYVHNGRFKEFLH 192
QY 170 AF-----HSVAFPS---C-----GPCLEQODEERSTR 193
DB 193 PYAESLVSGIGHVQELHRSVAPHAPASPARLSRCVQLSRKLTAKAKLHARIQNLQD 252
QY 194 LKDLIGRYFEESGQEPGEMDAQEPGQARLQDMEDVRCDIQF 240
DB 253 LRBELSRFAFAGT-----GTEBGAGPDP-----QMLSEVRORLQAF 288
```

RESULT 12
US-11-000-463-425

```
Sequence 425, Application US/11000463
Publication No. US20050266423A1
GENERAL INFORMATION:
APPLICANT: Tang, Y Tom
APPLICANT: Liu, Chenghua
APPLICANT: Aundul, Vinod
APPLICANT: Chen, Rui-hong
APPLICANT: Qian, Xiaohong B.
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Zhou, Ping
APPLICANT: Cao, Yi-Cheng
APPLICANT: Drmanac, Radoje T.
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 785CIP4CN
CURRENT APPLICATION NUMBER: US/11/000,463
CURRENT FILING DATE: 2004-11-29
PRIOR APPLICATION NUMBER: 10/291,265
PRIOR FILING DATE: 2002-11-08
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
```

```
PRIOR APPLICATION NUMBER: 09/922,279
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: 09/617,746
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/631,451
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 09/633,870
PRIOR FILING DATE: 2000-09-15
NUMBER OF SEQ ID NOS: 944
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 425
LENGTH: 400
TYPE: PRT
ORGANISM: Homo sapiens
US-11-000-463-425
```

```
Query Match 4.6%; Score 75; DB 7; Length 400;
Best Local Similarity 20.3%; Pred. No. 5.2;
Matches 46; Conservative 35; Mismatches 86; Indels 60; Gaps 8;
```

```
QY 54 AHCPCPLAVCLAQQLPEPDCGSSSVFEDMKLVDSMGHELASVRALCOLQMDHEPRFG 113
DB 116 APRLPQDPVGMKRQQLQEBLEEVKARLOPYMAEAEHVLGMNLGLRQQLKPYTMDLMEQVA 175
QY 114 VARGTGVLPSEFLAFHLRSPGDLTAEE---KDQICDFLYGRVOARERQALRLRTFQ 169
DB 176 LR-----VOELQELRVGSDTKAQLGVDAMALLQG-LQSRVYVHNGRFKEFLH 226
QY 170 AF-----HSVAFPS---C-----GPCLEQODEERSTR 193
DB 227 PYAESLVSGIGHVQELHRSVAPHAPASPARLSRCVQLSRKLTAKAKLHARIQNLQD 286
QY 194 LKDLIGRYFEESGQEPGEMDAQEPGQARLQDMEDVRCDIQF 240
DB 287 LRBELSRFAFAGT-----GTEBGAGPDP-----QMLSEVRORLQAF 322
```

RESULT 13

```
US-10-131-826A-136
Sequence 136, Application US/10131826A
Publication No. US20050245730A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Naureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C128
CURRENT APPLICATION NUMBER: US/10/131,826A
CURRENT FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
```

```

; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 136
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-136
```

```

Query Match      4.6%; Score 74; DB 6; Length 278;
Best Local Similarity 22.5%; Pred. No. 4.1;
Matches 64; Conservative 34; Mismatches 86; Indels 100; Gaps 15;
```

```

QY 9 MPEEIEITLCYELHHPHMLLELATYTHGRLNCPG--GPAQLOA-----LAHRCPL 60
DB 4 MPEEASRLCL---LLEPLLLLL---LLEPELSPQAGAEENDWVRLPEKC--- 49
QY 61 AVC--LAQQLP---EDPG-----QGSSEVPEFDMVKLVDSMGWELASVRAL 101
DB 50 EVCKYVAVELKSAPEEIKTEKVEIGTGYILDQKASGVKTKSDJ-----RLIEVTETI 103
QY 102 COLQMD--HEPRG-----VRGTVLVER-----SELAFHL 131
DB 104 CKRLIDYSILHKERTGSGNRFPAKGMSETFETLHNLVHKGVKVVMDIPLYELMNETSAEVA-DL 162
QY 132 RSPGDLTAEEBDOQCDPLFYGRVQAREQALARLRTPQAFHSVAPSPGCPCLSEQ----- 186
DB 163 KKQCDLVVEBEVEYIEDWYRNHQBED-----LTFELCANHVLMKXQDTSCLAEQMSGKK 215
QY 187 -----DEERSTRLKDLIGRYFEEBEGOEPCGMEDAQPE 220
DB 216 GDTAALGGKSKKSSRAKAGGSSSSSKQKKEGLGEGDPSPE 259
```

```

RESULT 14
US-10-821-234-1658
; Sequence 1658, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Precipita
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: PC_SEQ_genes Version 1.0
; SEQ ID NO 1658
; LENGTH: 777
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1658
```

```

Query Match      4.6%; Score 74; DB 6; Length 777;
Best Local Similarity 23.7%; Pred. No. 15;
Matches 67; Conservative 35; Mismatches 93; Indels 88; Gaps 16;
```

```

QY 11 BEAIEITLCYELHHPHMLLELATYTHGRLNCPGPAOL-QALAHRC 57
DB 157 EVAISVSTWLASHPEDFSGAKQOLDRLESFLLQTGAACK-GVGGASALINLRSRV 215
QY 58 PPLAVCLAQ--QLBEDPGQSSSVPEFDMVKLVDSMGWEL-----ASVRALCOLQ-----W 106
DB 216 DQAPDLPEKPLALGDDPADPTDV---LVPLADHIAQLTLLDABLEPLNLIPSCLGSLW 272
QY 107 DHEPRTG-----VRGTVLVEFSBLAFHLS-----PGDLT----- 138
DB 273 GHRDRPGHSHLCPGVR---ATVTQFNKVAAGAVSVLGATGTGCGPGEVTIRPLRPQRA 329
QY 139 -----AEKQICDF--LYGRVQAREQALARLR-----TFQAFHSVAPSPC 179
DB 330 RLEKMWIRVAECCLLNPFSSVAVVASALQSSPIHRLRAWGEATRDSLRVSSSL----- 384
QY 180 GPCLSEQDEERSTRLKDLIGRYFEEBEGOEPCGMEDAQPEPG 222
DB 385 --CQIFSEEDVYSQSRELL---VQSVLKQSLPERHSKAPRSG 422
```

```

RESULT 15
US-10-995-561-773
; Sequence 773, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: Cargill, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; PRIOR FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 773
; LENGTH: 3803
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-773
```

```

Query Match      4.6%; Score 74; DB 6; Length 3803;
Best Local Similarity 22.2%; Pred. No. 1e+02;
Matches 48; Conservative 34; Mismatches 72; Indels 62; Gaps 10;
```

```

QY 52 ALAHRCPPLAVCLAQQLPEDEPGQGSSEVPEFDMVKLVDSMGWELASVRALCOLQMDHEPR 111
DB 2032 SLAERSLSLOKAIQ-----SQSVQESLESLSIG-----EVEQNLGK 2071
QY 112 TGVRRGTGVLVEFSBLAFHLSPGDLTAER-----KDQICDPLFYGRVQAREQALARLR 165
DB 2072 QVSSLSGGVIOE--ALATNMKLQDIARQKSLERATREMTYRFMETADSTTAVALQGLA 2129
QY 166 RTPQAFHSVAPSPGCPCLSEQDEERSTRKDL--GRYFEEBEGOEPCGMEDAQ-----AOG 218
DB 2130 EVSQRFEQL-----CLQQDEKSS--LKXKLPLQAMFEHLSGKLQOFENKSRMLASG 2180
QY 219 PEPQARLQWEDQVRCDIROF-----LSLRPEK 248
DB 2181 NQPDQ-----DTHFPQIOELNEMEDQ 2204
```

```

Search completed: December 27, 2005, 22:12:55
Job time : 4 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 27, 2005, 21:40:26 ; Search time 45.6 Seconds
(without alignments)
2548.900 Million cell updates/sec

Title: US-09-889-325-4
Perfect score: 6424

Sequence: 1 MERLRDVERLRQAWERAFRR.....HLSPHALVGLATBELLOVAR 1208

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	872	13.6	941	2	A86404
2	631.5	9.8	610	1	BVECRQ
3	631.5	9.8	611	2	F86069
4	631.5	9.8	611	2	H91222
5	609.5	9.5	609	2	AH0917
6	593	9.2	610	2	A10466
7	581.5	9.1	620	2	E82351
8	560.5	8.7	602	2	H97265
9	560.5	8.7	602	2	AH2583
10	552	8.6	590	2	AD1419
11	551	8.6	718	2	AE1832
12	548	8.5	590	2	AE1794
13	545	8.5	712	2	E83226
14	542	8.4	496	2	A69691
15	525.5	8.2	1338	2	S62467
16	523	8.1	824	2	G75413
17	521	8.1	615	2	G87678
18	514	8.0	593	2	B89844
19	513	8.0	592	2	H86850
20	510.5	7.9	676	2	T34609
21	507	7.9	480	2	AB2411
22	506.5	7.9	809	2	T20430
23	505	7.9	1231	2	T24415
24	505.5	7.9	478	2	S77358
25	505.5	7.9	591	2	F69901
26	504	7.8	766	2	H81794
27	499	7.8	491	2	G83850
28	499	7.8	766	2	G81216
29	496.5	7.7	1436	2	T14895

30	493.5	7.7	645	2	A82689	DNA helicase XF138
31	489	7.6	1447	2	S50918	DNA helicase TP51
32	484	7.5	1401	2	T30247	DNA helicase PR
33	479.5	7.5	1401	2	T17452	Werner syndrome pr
34	478.5	7.4	714	2	G97230	recQ protein, supe
35	476	7.4	665	2	B86243	DNA helicase homol
36	464.5	7.2	659	1	A55311	DNA helicase RECQL
37	463.5	7.2	1417	2	A57570	Bloom's syndrome r
38	458.5	7.1	459	2	B89927	hypothetical prote
39	443	6.9	1031	2	G96634	probable DNA helic
40	437	6.8	467	2	AF1317	ATP-dependent DNA
41	434.5	6.8	467	2	AF1689	ATP-dependent DNA
42	432	6.7	886	2	T16536	hypothetical prote
43	422	6.6	483	2	T51906	related to recQ ge
44	405.5	6.3	607	1	C71367	probable ATP-depen
45	386	6.0	1273	2	E72611	probable ATP-depen

ALIGNMENTS

RESULT 1

probable protein ATP-dependent DNA helicase RecQ [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86404
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federgraj, N.A.; Kaul, S.; White, O.; Alonso, C.H.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulst, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; A.; Li, J.H.; Li, Y.; Liu, X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.; ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86404
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-941 <STO>
A:Cross-references: UNIPARC:UPI000017A6AC; GB:AE005172; NID:g10998928; PIDN:AAG26068.1; C:Genetics:
A:Map position: 1

Query Match 13.6%; Score 872; DB 2; Length 941;

Best Local Similarity 26.2%; Pred. No. 86-39;
Matches 305; Conservative 165; Mismatches 390; Indels 306; Gaps 43;

QY	93	TGGRSRQGVPPYGGRLKNTLQAGPALGRPPPLGRASSKASTPKPGTPVPSFA	152
DB	15	TPPRD---SPSPSPDLOSPAKHV---PVSRRK---TSSSSSKPKAP---THPPNFS	62
QY	153	EKVSDEPQLPPPPRPGRLQHLQASLSQRLSLPGLQRCSEVPDFIAPKACRPDL	212
DB	63	QAPVPSPFPPEPPPEP-----LFTNLPRRIGOSQA-----RSSVSST---SLCSRAS	111
QY	213	GSEESQLLPGESAVLPGAGSQGPESAFQEVSTRVSGPPSSSGGEXRNNNEPWSF	272
DB	112	FISVEKLKSDGVDFV-----PEPPLVEVIA-----	136
QY	273	AQVQSSSQAGPSSBAGAVAYEEDPRG---EPVQAQPPQPCSSPENPRYHGLSPSSQAR	329
DB	137	-----PPK-----SVRRKPPNLTITITISPPKPVNFRSN-----	166
QY	330	AGKAEGLTAHLHIFPRLARHGRGVYVFLNKKOKHYVGRALRSRL-----LRQAKQKX	383
DB	167	-GNGE-----GNFVTLNNGK---RGKFKP6KYKVSRSRSYSFRGR	206
QY	384	RKKGECPGGGATVTTKESCFLENGFDHMAACCPRASSEDIDAVPEPLVPSPPQVPEV	443
DB	207	YKKKEADGGESIL--EESSDLOKQI-----EDEANGF-----I	238

```

Qy 444 PSIDPTVLPVLSLPSGQLAETPAEVPAQLBQLGHQARRPGQERAVWMLISGISTLVLP 503
Db 239 SSVSDALIAVTEASDENLTKLNLVY-----GYDSFRDQLOAIRKILGSSSTMVLP 292
Qy 504 TGAKSLCYQCPALVYSRSPCLTVLVPPLSLMDQVSGLPCCAKACIHSGMTKORE 563
Db 293 TGAKSLCYQCPAMIL-----PGITLVSPVLSMLDQKHLPSITKIGGLSSSQAREAT 348
Qy 564 SVLQKIRPAOVHMLTPREALVGAAGLPAAQLPPVAFACIDEAHCLISQ----- 612
Db 349 ETLKRLKEGILKIVLFVSPERLNVFSLMFMSLSVSLVWVDEAHVSECICLSLFLCS 408
Qy 613 W-----SHNRPQCYLRV-CYKLRKMGVHCPLGLTATRTKRASVAVAGHLVAEAPD 663
Db 409 WLPALIRNSHNPSPSYRLKASMLPSELKAECLIAMATATMTLQAVMSLEI-PSGN 467
Qy 664 LAGPAPVPTNLHLVSM---DRPTDQALLTLQGRFQNLISIIYCNRRREDTERIALTL 720
Db 468 LIQSKQARDNFELSVLSGARNMD--LLILMSPPYEISIIYVCKFOYETDMISKYL 525
Qy 721 RTCLHAAMPVSGSGAPRTTAEAYHAGNCSRRRRVQRAFMQQLRVVATVAFQMGILDR 780
Db 526 RD-----NNINAKGYHSGLPKXDRVRIQESFCSNKRIVVATVAFQMGILDX 571
Qy 781 PDVAVNLHLGLPSPFESYVQVAGARGDQPAHCHLPLOQGEDELRLRHVADSTDFL 840
Db 572 GDVAIVHIFSVPGMEEVQEIIGRARGRLSYCHLFY--DNDTYLTKRLSHSGVDEY 629
Qy 841 AVKLVORVPFACCTCTPRPSBQSGAVGSRPVYKTPQBAQLSHOAPGPRVCGH 900
Db 630 AVGKFLTHVF-----STETKQHE-----KIC--- 650
Qy 901 ERALPIQLTVQALMPBEALITLLCYELAPHHMLLELTATYTHGRNLCPGPAQLQALA 960
Db 651 --SLVBSASQKPFMKKEVMQITILHLEIGVQYLRMLPOLNICTTAN-----F 697
Qy 961 HR---CERPLAVCLAQQLPBDGCGSSSVYEFDMVTKLVDSMG---MELASVRALCOLQMDH 1014
Db 698 HKTSDAQNTVWLYVOSSFNTLAARSATVAALIKSHVQGLHVEDIPVAVASICAT--- 754
Qy 1015 EPRTVGRGTGLVLEFSELAFLHRSPLDTAAEKD-----QICDFLGRVQAR 1062
Db 755 -----IDVLAELQAL-----KGVVTELDQSAFCYITILKSPKICLSLHITKWL 799
Qy 1063 ERQALALRR-TFOAFHSVAFPSGCPCLGQDEERSTLKDLLGRYFEEBEGQEPGQMED 1121
Db 800 TEIISKVKRKLIDIMSSAAVAIVSNTSELSSGAKQTR--SLQSRIFDYFNG-----D 850
Qy 1122 AQGEBEPQARLQWEDQ---VRCDIRQFLSLRBEKSSRAVARIFHGIGSPCYPAQVY 1177
Db 851 EKCDSPSKA-----TQWCAFLRADIKVFLQSNRQAKETPRAIARIMGVSPAFPMVSW 904
Qy 1178 GQDRFRFRKYLHLSPHALVGLATEEL 1203
Db 905 SK-THFNGRIYAVDFRVIMEAAQTEL 929

```

RESULT 2

```

BVECRQ
DNA helicase recQ - Escherichia coli (strain K-12)
N/Alternate names: DNA-dependent ATPase recQ (EC 3.6.1.-)
C/Species: Escherichia coli
C/Date: 31-Dec-1989 #sequence, revision 10-Oct-1997 #text_change 01-Mar-2002
C/Accession: G65186; J05137; A55776; S30712
R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426617; PMID:9278503
A/Accession: G65186
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA

```

```

A/Residues: 1-610 <BLAT>
A/Cross-references: UNIPARC:UPI00001661F2; GB:AE000458; GB:U00096; NID:92367299; PIDN:AAC
A/Experimental source: strain K-12, substrain MG1655
A/Ritino, N.; Nakayama, K.; Nakayama, H.
Mol. Gen. Genet. 205, 298-304, 1996
A/Title: The recQ gene of Escherichia coli K-12: primary structure and evidence for SOS
A/Reference number: J05137; MUID:87115164; PMID:3027506
A/Accession: J05137
A/Molecule type: DNA
A/Residues: 1-256, 'A', 258-610 <IR1>
A/Cross-references: UNIPARC:UPI000016F68A
A/Experimental source: strain K12
A/Note: it is uncertain whether residue 1, 3, 8 (all coded by GTG), or 9 (coded by TTG)
A/Unzu, K.; Nakayama, K.; Nakayama, H.
Proc. Natl. Acad. Sci. U.S.A. 87, 5363-5367, 1990
A/Title: Escherichia coli recQ protein is a DNA helicase.
A/Reference number: A35776; MUID:90319113; PMID:2164680
A/Accession: A35776
A/Status: preliminary
A/Molecule type: protein
A/Residues: 4-8 <UMB>
A/Cross-references: UNIPARC:UPI0000172B8C
A/Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A/Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to
A/Reference number: S30660; MUID:92358234; PMID:1379743
A/Accession: S30712
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 'V', 2-256, 'A', 258-610 <DNA>
A/Cross-references: UNIPARC:UPI0000172EBD; EMBL:M87049
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1992
A/Genetics:
A/Genes: recQ
A/Map position: 85 min
A/Start codon: GTG
C/Function:
A/Description: involved in the recQ recombination pathway; its gene expression is under
A/Pathway: recQ recombination
C/Superfamily: recQ protein, recQ helicase homology
C/Keywords: ATP; DNA binding; hydrolase; nucleotide binding; P-loop; recQ recombination
C/49-56/Region: nucleotide-binding motif A (P-loop)
C/144-149/Region: nucleotide-binding motif B
C/146-151/Region: DNA motif
C/369-402/Domain: recQ helicase homology <RHH>
Query Match 9.8%; Score 631.5; DB 1; Length 610;
Best Local Similarity 37.6%; Pred. No. 2.8e-26;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;
Qy 474 EQLGHQARRPGQERAVWMLISGISTLVLPPTGAGKSLCYQLPALVYSRSPCLTVVSP 533
Db 21 ETFGQGFRRQGEIITDVTLSGRDCLVMPPTGGKSLCYQIPALLLNG-----LTVVSP 76
Qy 534 LSLMDQVSGLP-CLKAACIHSGMTKQRESVLQKTRAAQVHVMMLTPRALVAGAGLPP 592
Db 77 ISLMDQVDPQLQANGVAAACINSGTQREQLLEVWTCGRTQIRLLTYIAPRLMDNPLN 136
Qy 593 AAOQLPVAFCIDPEAHCTLSQMSHNPFCYLRVCKVLEERGNGVHCPLGTATATRTASDV 652
Db 137 LAHNPVLLA-VDEAHCTLSQMSHNPFCYLRVCKVLEERGNGVHCPLGTATATRTASDV 194
Qy 653 AOHVAEAPDGLGAPVPTNLHLVSMDBRDTQALLTLQGRFQNLISIIYCNRRREDTERIA 703
Db 195 VRLIG-----LNDPL-----IQIS-SFDRPNRYMLM-----EKFKLDDQMRVVGORG 238
Qy 704 ---IYCNRRREDTERIALIRTCILHAAMPVSGSGARPKTTAEAYHAGNCSRRRRVQRAF 760
Db 239 KSGIYCNRSRAKVEDTAAIR-----QSKGISAAAYVHAGLENNVADVQEK 284
Qy 761 MGOGLRVVAVTVAFGWGLDPDVRVAVLHLGLPSPFESYVQVAGARGDQPAHCHLFL-- 818
Db 285 QRDLDQIVAVTVAFGWGLDPDVRVAVLHLGLPSPFESYVQVAGARGDQPAHCHLFL-- 344

```

QY 819 -----OPOGEDLRELRH---VHADSTDFLAVKRLV-----QKVPACTC- 855
DB 345 ADMAMLRCLBEKPGQ--LQDIERHKLAMGAFBAQTCRLVLNLNYGSGRQEPFCGNC 403
QY 856 TCTRPSEOGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 3

ATP-dependent DNA helicase [imported] - Escherichia coli (strain O157:H7, substrain EDLg)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: F60693
R/Perna, N.V.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: AB5480; MUID:21074935; PMID:11206551
A/Accession: F6069
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-611 <STO>
A/Cross-references: UNIPROT:Q8X8N1; UNIPROT:Q8FBM6; UNIPARC:UPI00000D08FF; GB:AB005174;
A/Experimental source: strain O157:H7, substrain EDLg33
C/Genetics:
A/Gene: recQ
C/Superfamily: recQ protein; recQ helicase homology

Query Match 9.8%; Score 631.5; DB 2; Length 611;
Best Local Similarity 37.6%; Pred. No. 2.8e-26;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;
QY 474 BOLGHOARPGQERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPCLTLVSP 533
DB 21 EFTGIGQPRPGQEBITDVLGSDCLVMPFTGGKSLCYQIPALLNG---LTVVSP 76
QY 534 ISLMDQVSGLP-CLKAACHSGMTRKQRESVLQKIRAAQVHMLTPBALVGAGLPP 592
DB 77 ISLMDQVQLOANGVAACLNSTQTRQLEVMTCGTGQIRLLYIAPERLMDNLFLEH 136
QY 593 AAOLEPPVAFACIDEAHCLSGSHNRPCTLYKCVLRBMGVHCLGATATRTASDV 652
DB 137 LAHNPVLLA-VDEAHCLSGMTRKQRESVLQKIRAAQVHMLTPBALVGAGLPP 194
QY 653 AAOLEPPVAFACIDEAHCLSGSHNRPCTLYKCVLRBMGVHCLGATATRTASDV 703
DB 195 VRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM---EKFKPDQMLKRYQEOG 238
QY 704 ---IYCNREDTERIALLRCTLAHAAWPGSGRAA-PKTTAAVYHAGKMSRRRRVQRAF 760
DB 239 KSGIITCNSRAKVEDTAARL-----QSKGISAAAYHAGLENNVADVQEKF 284
QY 761 MGOGLAVVAVAFMGGLDRPDVAVLHLGLPSPSESVQAVGRGQCPAHCHFL-- 818
DB 285 QRRDLOIVAVAFMGGLDRPDVAVLHLGLPSPSESVQAVGRGQCPAHCHFL-- 818
QY 819 -----QPOGEDLRELRH---VHADSTDFLAVKRLV-----QKVPACTC- 855
DB 345 ADMAMLRCLBEKPGQ--LQDIERHKLAMGAFBAQTCRLVLNLNYGSGRQEPFCGNC 403
QY 856 TCTRPSEOGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 4

H91222
ATP-dependent DNA helicase [imported] - Escherichia coli (strain O157:H7, substrain RIMD)
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C/Accession: H91222
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Ganeara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A/Reference number: A96629; MUID:21156231; PMID:11258796
A/Accession: H91222
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-611 <HAV>
A/Cross-references: UNIPROT:Q8X8N1; UNIPROT:Q8FBM6; UNIPARC:UPI00000D08FF; GB:BA000007; I
A/Experimental source: strain O157:H7, substrain RIMD 050952
C/Genetics:
A/Gene: ECE4752
C/Superfamily: recQ protein; recQ helicase homology

Query Match 9.8%; Score 631.5; DB 2; Length 611;
Best Local Similarity 37.6%; Pred. No. 2.8e-26;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;
QY 474 BOLGHOARPGQERAVMRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPCLTLVSP 533
DB 21 EFTGIGQPRPGQEBITDVLGSDCLVMPFTGGKSLCYQIPALLNG---LTVVSP 76
QY 534 ISLMDQVSGLP-CLKAACHSGMTRKQRESVLQKIRAAQVHMLTPBALVGAGLPP 592
DB 77 ISLMDQVQLOANGVAACLNSTQTRQLEVMTCGTGQIRLLYIAPERLMDNLFLEH 136
QY 593 AAOLEPPVAFACIDEAHCLSGSHNRPCTLYKCVLRBMGVHCLGATATRTASDV 652
DB 137 LAHNPVLLA-VDEAHCLSGMTRKQRESVLQKIRAAQVHMLTPBALVGAGLPP 194
QY 653 AAOLEPPVAFACIDEAHCLSGSHNRPCTLYKCVLRBMGVHCLGATATRTASDV 703
DB 195 VRLLG-----LNDPL-----IQIS-SFDRPNIRYMLM---EKFKPDQMLKRYQEOG 238
QY 704 ---IYCNREDTERIALLRCTLAHAAWPGSGRAA-PKTTAAVYHAGKMSRRRRVQRAF 760
DB 239 KSGIITCNSRAKVEDTAARL-----QSKGISAAAYHAGLENNVADVQEKF 284
QY 761 MGOGLAVVAVAFMGGLDRPDVAVLHLGLPSPSESVQAVGRGQCPAHCHFL-- 818
DB 285 QRRDLOIVAVAFMGGLDRPDVAVLHLGLPSPSESVQAVGRGQCPAHCHFL-- 818
QY 819 -----QPOGEDLRELRH---VHADSTDFLAVKRLV-----QKVPACTC- 855
DB 345 ADMAMLRCLBEKPGQ--LQDIERHKLAMGAFBAQTCRLVLNLNYGSGRQEPFCGNC 403
QY 856 TCTRPSEOGAVGGE 871
DB 404 ICLDPPKQYDGSSTDAQ 419

RESULT 5

AH0917
ATP-dependent DNA helicase (BC 3.6.1.-) [imported] - Salmonella enterica subsp. enterica
C/Species: Salmonella enterica subsp. enterica serovar Typhi
A/Note: This species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AH0917
R/Farhili, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moutle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A/Authors: Farry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AH0917
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-609 <PAR>
A/Cross-references: UNIPARC:UPI000005A56B; GB:AL513382; PIDN:CAD07934.1; PID:g16504479; C
C/Genetics:

A:Gene: recQ
C:Superfamily: recQ protein; recQ helicase homology
C:Keywords: hydrolase

Query Match	9.5%	Score 609.5;	DB 2;	Length 609;
Best Local Similarity	37.7%	Pred. No. 4.1e-25;		
Matches 162;	Conservative 58;	Mismatches 137;	Indels 73;	Gaps 14;

```

QY 474 QOLHQAEPPOEBAVMHILSGISTLLVLPNGAGSLCYOLPALLYSRRSPCLTIVVSP 533
Db 19 ETFYQOPPOEALITPALSGRDLVMPFGGKSLCYQIPALLDG---LTVVSP 74
QY 534 LSLMDVYSG-LPPLCAACIHSGMTKORSBVLQIKRAQVHVMILTEALVGAGLPP 592
Db 75 ISLKKQDQDLANGVAACILNSTQSRQGLEVMAGCMTGQIRLLIYAPERLMDNFDH 134
QY 593 AAQLEPFAACIDEAHCLSSOWSHNRPCTYLRVCYLREMGVHCLGLTATITRTTASDV 652
Db 135 LAHNPVLLA-VDEAHCTISOWGHDFRPEYALGQ-LRORFALPMLATATADTTRODI 192
QY 653 AOHFAVAEPPDLHGAPVPTMILHSVSMRDPDOLATLLQKRONDSI----- 703
Db 193 IRLLG-----LNDPL-----IQIS-SFDRPIRITLM---EKFKPDLQRLRYQEOBG 236
QY 704 ---IYCNREDTERIALLRCTLHAAVWPSSGGAAPTAEAYHAGMSCRRERREVOBAF 760
Db 237 KSGIITCSNRKAVEDTARLQ-----SRG-----ISAAVYHAGLENAIRADVQKF 282
QY 761 MOGQLVAVVAIVARBMGLDPDVAVVLHLGPRSESYQAVGARGRGQPAHCHLFIQ 820
Db 283 QORDLIQVAVTAVAFMGINKPRVRFVHPDI-PRNIESYVQELGGRAGDGLPEAMLFDP 342
QY 821 -----QGEDLRELRHVAHUSTDIAVRLV-----QRVPACTG-T 856
Db 343 ADMAMLRRCLEBKPAQGLQDIERHHTLNMGAFALEQTCRLVLLVYFPGEGQEPGCDI 402
QY 857 CTRPPEBOEG 866
Db 403 CLDPKQYDYG 412

```

RESULT 6
A10466
ATP-dependent DNA helicase (EC 3.6.1.-) [imported] - *Yersinia pestis* (strain CO92)

C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: A10466
P:Perkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; PMID:1156360
A/Accession: A10466
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-610 <KUR>

A/Cross-references: UNIPROT:Q8ZAG8; UNIPARC:UP10000DDCAD; GB:AL590842; PIDD:CA93301.1
C/Genetics:
A/Gene: recQ
C/Superfamily: recQ protein; recQ helicase homology
C/Keywords: hydrolase

Query Match	9.2%	Score 593	DB 2	Length 610
Best Local Similarly	33.3%	Pred. No. 3.1e-24		
Matches 169	Conservative	71	Mismatches 172	Indels 96
				Gaps 18

QY 474 BOLGHOAFRRGERAVRMILISGISTLLVPTGAGSLCYOLPALYRSRSPCLTLYVSP 533
 QY 19 DTFGYGQFRNGQGEIINATLSGSDCLVMPFTGGSLCYQIPALV----TGTGLTVVSP 74
 QY 534 LSLMDQVSGSLPPC-LKAACTHSGMTRKQRESVTLQIRAAQVHVIMLTPEALVGAGGILPP 5929

```

Db      75 ISLMKQVDQLLAVGVAGCINSSQTRBQQLAVMDGCRGQIKLLYLA PERLVNESFLDQ 134
Qy      593 AAOLPVAFCIDBAHCISOWSHNFRPCYLEVCVKEIRBMGVHCFGLTATATRTASDV 652
Db      135 LYQWRP-ALLAVDEAHCTISQMGHDFRBEY-RALCOLKORFPDLFVIALTATADENATRGDI 192
Qy      653 AQHILVAEERPDLAGPAPVPTMILSVSMEDRPTDALLTLQSKRFQUNDISI----- 703
Db      193 VRLL-----NLDGPL-----IQIS-SFDRNITYTLV----EKFKPLDQIMRFPVQDORG 236
Qy      704 ---IYCNRREDTERIALLTCTCLHAAMVPGSGGRAPKTTAAEAYHAGMCSNERRRVORAF 760
Db      237 KSGIITVCSRAKVEDTJARLQ-----SRG-----LSVAAYHAGLDNRRRAQVGEAF 282
Qy      761 MGGQLRVVATVAFQMGGLDRPDVAVLHLGLPSPFSZYVQAVGABAGDQORAHCHLFLQRP 820
Db      283 QRDQLQVVAVAVAGMGINKKNVAFVVFHFDLPRTIESYQETGABAGDGLPAEAVLLYDRP 342
Qy      821 QG-----EDIRELRHHVHADSTDFLAVKRLVORVF-----DACTC-T 856
Db      343 ADMAMLRRCLEEKPAQAGAOCTIERHKLNAMGAFABAOCTCRRLVLLNTVFOEGKQOCGNCIDI 402
Qy      857 CTRPPSEDEGAVGGERPVP-KYRPQEZABQJSH-----QABGPRRVCKGHERALPIQJLTV 910
Db      403 CLDEPKRVDGLADAOXKALSCVYRVGORFGJGYIVEVLRCANNOIRIREWGHDK-----LSV 457
Qy      911 QALMDPEBAITETLLCYLELHPHMLLELL 938
Db      458 YGIGR-----EQTEHHWVSVL 473

```

```

RESULT 7
E82351
AMP-dependent DNA helicase RecQ VC0196 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

```

C:Accession: E82351
R:Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, J.
L.R.; Mekalanos, J.J.; Venter, J.C.; Frazer, C.M.

Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833; PMID:1095301

A:Accession: B82351
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-520 <HEI>
A:CDSs-references: UNIPROT:Q9KYF0, UNIPARC:UPI00000C2C1C, GB:AE004110, GB:AE003852, NIT
A:Experimental source: serogroup O1, strain N16961, biotype EI TOR
C:Genetics:

A;Gene: VC0196
A;Map position: 1
C;Superfamily: recQ protein; recQ helicase homology

Query Match	9.1%	Score 581.5;	DB 2;	Length 620;
Best Local Similarity	33.0%	Pred. No. 1.3e-23;		
Matches 160; Conservative	74;	Mismatches 171;	Indels 80;	Gaps 14

[illegible]

695 KRPNLDSIIICNRREDTERIALRTCLHAAMVWGSGRAPKTTAAVHAGMSRRR 754
246 QRCQ---CGIIYCSRKVENLTETL-C-----GNHIR-----AASYHAGMDADRA 288
755 RVORAFMGOGLRVVATVAFMGIDRDPVRAVLHLGLPPSFESYQAVAGARDGQPAHC 814
289 WVQBAFQGDLDQIVATVAFMGINKPNVRFVHFDPRIANISYQETGAGRDGLPABA 348
815 HLFLOPQGEDELRELRHHVHADSTDFLAKYLVQVFPACTCTCTPRPSEOGAVGGERPV 874
349 MMLYDP--ADNMWLRRLM-----DEKPD 369
875 PKYRQAEQSLSHQAAPGPRVC-----MGERALPIQLTVOALDMP-----BEAL 920
370 GAQKQVESHKLTAMGAFABAQCTCRQVLTNFTGEYTRDXPCGNCICLDPPKRFDAETEAR 429
921 ETLIC 925
430 KALSC 434

RESULT 8

H97365
DNA helicase Xf1381 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: H97365
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H97365
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-602 <KUR>
A:Cross-references: UNIPROT:Q8U77; UNIPARC:UP100000D16A7; GB:AE007869; PIDD:AAK85881.1; C:Genetics:

A:Gene: AGR_C_92
A:Map position: circular chromosome
C:Superfamily: recQ protein; recQ helicase homology
Query Match 8.7%; Score 560.5; DB 2; Length 602;
Best Local Similarity 34.6%; Pred. No. 1.6e-22;
Matches 151; Conservative 73; Mismatches 152; Indels 61; Gaps 17;

466 PAVFOALQOLGHOAFRPGQERAVNRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPC 525
5 PLNLIKAV--YGYDAFRGQGEIIQHVVAAGNNAFVLMPTGGKSLCYQIPAL--AREG-- 58
526 LTLVSPPLSLMDQVSGLPFC-LKAACTHSGMTRKQRESVLQKTRAAOVHVMLTPEAL 584
59 MGLVSPPLALMWDQVVALRQAGVABALNSDLSPEERTLMRDVAGNVLDLYAAPETL 118
585 VAGAGLPAAQUPVAFACIDEAHCLSQMSHNPFCYLRVCKVLRBMGVHCFGLTATA 644
119 LKPDVL-DALQSDISLVAIDEAHCLSQMGHDFRPY-RQDMLARFPNTPRMALITATA 176
645 TRRTADVAOHLVAEEDPLHGPAVPNTMLHSVMDRDTQALLTLQGRFON---LD 701
177 DEPTAELIIGHLAI-DEADAFIAGFDRPNIRAI-MEKDNPRTOJ-----KFLNGRENE 229
702 SIIICNRREDTERIALRTCLHAAMVWGSGRAPKTTAAVHAGMSRRRRVQRAFM 761
230 SGIVYCLSKRKVEETAAWLR-----EEGR---DALPYHAGMDKAAAEENQTRPQ 275
762 OGQLRVVATVAFMGIDRDPVRAVLHLGLPPSFESYQAVAGARDGQPAHC----- 814
276 HGEAVIIVATVAFMGIDRDPVRAVLHLGLPPSFESYQAVAGARDGQPAHC----- 814
815 -----HLFLOPQGEDELRELRHHVHADSTDFL-----AVKRLV-----QRVFPACTC- 855

Db 336 DIALNRPIEB--SDAADQKTRMERQKLDALLGLAETAGCRQVLTSTFGDRCEPCGNC 393
Qy 856 TCTRPSEOGAVGGER 872
Db 394 TCSSPDLFEAGATAOK 410

RESULT 9

AHP-dependent DNA helicase recQ [imported] - Agrobacterium tumefaciens (strain C58, Dupont AH2583)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AH2583
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Wang, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCellie, J.; Karp, P.; Romero, P.; Zhang, S. Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Kreespan, W.; Perry, M.; Gordon-Kamm, J.; Ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2583
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-602 <KUR>
A:Cross-references: UNIPROT:Q8U77; UNIPARC:UP100000D16A7; GB:AE008688; PIDD:AAL41086.1; A:Experimental source: strain C58 (Dupont)

C:Genetics:
A:Gene: recQ
A:Map position: circular chromosome
C:Superfamily: recQ protein; recQ helicase homology
Query Match 8.7%; Score 560.5; DB 2; Length 602;
Best Local Similarity 34.6%; Pred. No. 1.6e-22;
Matches 151; Conservative 73; Mismatches 152; Indels 61; Gaps 17;

466 PAVFOALQOLGHOAFRPGQERAVNRILSGISTLLVPTGAGKSLCYQLPALLYSRRSPC 525
5 PLNLIKAV--YGYDAFRGQGEIIQHVVAAGNNAFVLMPTGGKSLCYQIPAL--AREG-- 58
526 LTLVSPPLSLMDQVSGLPFC-LKAACTHSGMTRKQRESVLQKTRAAOVHVMLTPEAL 584
59 MGLVSPPLALMWDQVVALRQAGVABALNSDLSPEERTLMRDVAGNVLDLYAAPETL 118
585 VAGAGLPAAQUPVAFACIDEAHCLSQMSHNPFCYLRVCKVLRBMGVHCFGLTATA 644
119 LKPDVL-DALQSDISLVAIDEAHCLSQMGHDFRPY-RQDMLARFPNTPRMALITATA 176
645 TRRTADVAOHLVAEEDPLHGPAVPNTMLHSVMDRDTQALLTLQGRFON---LD 701
177 DEPTAELIIGHLAI-DEADAFIAGFDRPNIRAI-MEKDNPRTOJ-----KFLNGRENE 229
702 SIIICNRREDTERIALRTCLHAAMVWGSGRAPKTTAAVHAGMSRRRRVQRAFM 761
230 SGIVYCLSKRKVEETAAWLR-----EEGR---DALPYHAGMDKAAAEENQTRPQ 275
762 OGQLRVVATVAFMGIDRDPVRAVLHLGLPPSFESYQAVAGARDGQPAHC----- 814
276 HGEAVIIVATVAFMGIDRDPVRAVLHLGLPPSFESYQAVAGARDGQPAHC----- 814
815 -----HLFLOPQGEDELRELRHHVHADSTDFL-----AVKRLV-----QRVFPACTC- 855
336 DIALNRPIEB--SDAADQKTRMERQKLDALLGLAETAGCRQVLTSTFGDRCEPCGNC 393
856 TCTRPSEOGAVGGER 872
394 TCSSPDLFEAGATAOK 410

RESULT 10

AD1419
ATP-dependent DNA helicases homolog 1mo2757 [imported] - Listeria monocytogenes (strain 1
C:Species: Listeria monocytogenes

Db 67 ISLMKQDALVSDGIATFTNSTLTNRREIDIRLDAEFGELKMLYAPIERIETPGFORL 126
QY 593 AAQAPVAFACIDEAHCSQMSHNPFCYLRYCKYLBERMGVHCIGLTAATRTASDV 652
Db 127 IEQVPISTFA-IDEAHCISQMGHDFRPSYLSICDSLDMKTRRPVIALTAITQAVSDI 185
QY 653 AQHLAAVEEPDLHGPAVPPT-----NLHLVSMDRDTQALLTLLOGRFONLIIYYC 707
Db 186 CRLKI-KPN-----SVIKTGFSRNLAFOVVKGGDKKYLIDYLTG--NSTESGIYA 236
QY 708 NREEDTERIALRLCTLAAMVPGSGRAPKTTAA--YHAGMCRERRRRVRAWQOL 765
Db 237 STRKEVERHISFTL-----KKGVESGMVHGMTDIARDKDWQEKFLYDI 280
QY 766 RVVAVTVAFGKGLDPRDAVLHLGLPSPFESYVAVGGRAGDGPACHLFLQFQEDL 825
Db 291 RVIATVNAFGMGINKSNVAFVIHINIPIVTEATYQDAGAKHGVPSDCITLIFSPDSEKI 340
QY 826 RE 827
Db 341 QQ 342

RESULT 13

ATP-dependent DNA helicase RecQ PA3344 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: E83226
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lam,
; Loty, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83226
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-712 <STO>
A:Cross-references: UNIPROT:Q9HY01; UNIPARC:UPI00000C59C6; GB:AE004757; GB:AE004091; NID
A:Experimental source: strain PA01
C:Genetics:
A:Gene: recQ; PA3344
C:Superfamily: recQ protein; recQ helicase homology

Query Match 8.5%; Score 545; DB 2; Length 712;
Best Local Similarity 28.5%; Pred. No. 1.3e-21;
Matches 221; Conservative 102; Mismatches 294; Indels 158; Gaps 32;

QY 471 QALBOL-----GHQAPRPGQERAVMRILSGISTLVLPTGAGKSLCYOLPALYSRRSPCL 526
Db 4 QALRIKDVFGDAFRGNQRIIERVAEGGDALVLMPTGGKSLCYOLPALYREG---L 59
QY 527 TLVAVSLTSLMDQVS-----GLPCLKAKAICHSGMTRKQRESVLOKIPAAQVHVLMLTP 581
Db 60 TVVAVSLTSLMDQVATLDELGP-----AVALNSTINPEQORIALRLGRIKILYLP 115
QY 582 EALVAGGLPPAAQLPPVAFACIDEAHCSQMSHNPFCYLRYCKYLBERMGVHCIGL 641
Db 116 ERLVQPRMALFLQRLTFLVGLFA-IDEAHCISQMGHDFRPSYLSICDSLDMKTRRPVIALTAITQAVSDI 173
QY 642 ATATRTASDVQHLAAVEEPDLHGPAVPPTNLHLVSMDRDTQALLTLLOGRFONL 701
Db 174 ATDMETREMIQRLHLQNAEQFLSSFDPRNITRYIVPKQPKQ-LIGFLSRKQ--D 229
QY 702 SIIIVNREEDTERIALRLCTLAAMVPGSGRAPKTTAA--YHAGMCRERRRRVRAWQOL 761
Db 230 AGIVYCLSRKRVAVFL-----GNQ-----FRALPYHAGISNELRAHMQKFL 275
QY 762 QGGLRVVAVTVAFGKGLDPRDAVLHLGLPSPFESYVAVGGRAGDGPACHLFLQFQ 821
Db 276 NEBGLIMVATIAFGMGINKSNVAFVIHINIPIVTEATYQDAGAKHGVPSDCITLIFSPDSEKI 345

QY 822 GEDIRELRHVHADSTPLAVKRVQVFPACTCTTRPSEQGANG--GERPVP-- 875
Db 336 --DVILLRQMQSSECD-ERHKKRVERHLEAMLCERTRCRQALALVPEEMQPCGH 392
QY 876 -----KRPPEAQLSHQAP-----GPRVCMGHERALPIQLTVQALDMEBAIETLCC 925
Db 393 CDNCVDGVEVTDATESARQALSAIRSGQRYGVH-----LVDLIGRETETKRLS-- 443
QY 926 YLELHPH-----HW-----LELATYTHCRNCPGPAQLOALAHRCPP-- 966
Db 444 ---GHQLAVRGICKRGDEDEMRLLFRQLVARGLADVDLDFGG---LRLTEACRPLLR 496
QY 967 ---AVCLAQQLPEDPGQSS-----SVEFDYKLYDSKMGW-LASVYRALCOLQMDHERPT 1018
Db 497 GEVRLLELRDLKPPORAKSSSGGASASQLVRSERREWEALRLRLKLAE---EHSVVP 553
QY 1019 GVRKGTGVLVFSFLAFLRLSPDLTAEKQICDPLVGRVQAREQALARLRTFOAF- 1077
Db 554 YVIPPDATLLEM--LRQPRSLSDMAQ-----VSGVGAR-----KLERYQAF 595
QY 1078 -----HSVAFPSG-GRCLQODEERSTRKDLGRYFEEEGQEPGG 1118
Db 596 DVLTDSPAAPAPPODLRHLASLACAGMTPAQLARQNCSEKNVYMLAEALINQOVSL 655
QY 1119 MEDAGPPEPQARLQDWEQVRCIDTROPF-----SLRP-----EKKFSRAVARIFH 1165
Db 656 EQALDLPPELLIGEIQD-----AFLEBDGELPVAALBERGKRPVSGVLH 700

RESULT 14

ATP-dependent DNA helicase homolog - Bacillus subtilis
N:Alternate names: ATP-dependent DNA helicase recQ
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: A69691
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertore
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 380, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kunita, K.; Lapidus, A.; Lardinois,
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadleir, Y.; Sato, T.; Scanlon,
A:Authors: Scheich, S.; Schroeter, R.; Scorfione, F.; Sekiguchi, J.; Sekowska, A.; Sero,
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tepstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, B.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:198044033; PMID:9384377
A:Accession: A69691
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-496 <KUN>
A:Cross-references: UNIPROT:P50729; UNIPARC:UPI000006066D; GB:Z99115; GB:Z99116; GB:AL001
A:Experimental source: strain 168
C:Genetics:
A:Gene: recQ
C:Keywords: ATP; nucleotide binding; P-loop
F:38-45/Region: nucleotide-binding motif A (P-loop)
F:132-137/Region: nucleotide-binding motif B
F:136-139/Region: DEAH motif

Query Match 8.4%; Score 542; DB 2; Length 496;
Best Local Similarity 36.8%; Pred. No. 1.2e-21;
Matches 137; Conservative 68; Mismatches 117; Indels 50; Gaps 12;

QY 471 QALBOL-IGHQAPRPGQERAVMRILSGISTLVLPTGAGKSLCYOLPALYSRRSPCLTV 529
Db 6 QTLVQPFPGTSPKQGDIIISLGSKDTIAMLPTGGKSLCYOLPGYMLDG---WILL 61

```

QY      530 VSPILSLIMDDVSGIPPC--LKAACISIGGMRKORKESTLQKIRPAQAQVHVLMTGEALVGA 587
Ddb     62 VSPILSLIMEDVOQLKARGEKRAAANLNSMNRKOEQVLEHTIR--RYFETLYLSPALQSP 119
QY      588 GGLPPAALCPVPAACIDBAHCLSQWSHNPFCYLKVCVLREMGVHCIFGLTATATRR 647
Ddb     120 YLEKLSKV--PISLFEVIDEALHCISEMGHDFRPDSKLGQ--LRKLGHPPLVATLNTATKE 177
QY      648 TASYV-----AQHLVAABEPDLHGPAVLPYTNLHLSVSMNDRPTDQALLTLQGRFQ 698
Ddb     178 TLQGVNMLLELQHAVRHLSNVRR-----NIALRVENADTAEKIRVIO--LVE 225
QY      699 NLDSI-IIYCNRRREDTERIALLRCTCLHAAPVPGSGRAPPTTAA--YHAGMSRRERR 755
Ddb     226 NLQSPGVIYCTCRKMAKELAGEIKS-----KTSRPADYTHGSLSGDKRIL 270
QY      756 VQAPMGOQLRVVAVATVAFGKGLDRPDVRAVLHLGLPPSFESYVQAVAGRAGDQPAHCH 815
Ddb     271 IQQGFHINQGLDVICTNAFGMGVKXPDIRYVIHFFILPQTALAPMGEIGRARDKGPSTSI 330
QY      816 LFLQPGSEDLRE 827
Ddb     331 LLRAPGDFELQF 342

RESULT 15
S62467
ATP-dependent DNA helicase hus2 - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 16-May-1996 #sequence revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S62467; T38578; T43265
R/Badcock, K.; Churcher, C.M.
submitted to the EMBL Data Library, October 1995
A/Reference number: S62445
A/Accession: S62467
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-1328 <BAD>
A/Cross-references: UNIPROT:Q09811, UNIPARC:UPI000012CE81, EMBL:Z54354, NID:g1019398, PDB:1A98
R/Badcock, K.; Churcher, C.M.; Barrett, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A/Reference number: Z21745
A/Accession: T38578
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1328 <BA2>
A/Cross-references: UNIPARC:UPI000012CE81, EMBL:Z54354, PIDN:CAA91177.1, GSPDB:GN00066,
R/Stewart, E.; Chapman, C.R.; Al-Khodairy, F.; Carr, A.M.; Enoch, T.
EMBL J. 16 2682-2692, 1997
A/Title: Rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes,
A/Reference number: Z07997; MUID:97327548; PMID:9184215
A/Accession: T43265
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1328 <STE>
A/Cross-references: UNIPARC:UPI000012CE81, EMBL:Y09426, NID:g1684753, PIDN:CAA70577.1; F
C/Genetics:
A/Gene: hus2; SPAC2G11.12
A/Map position: 1L
C/Superfamily: Bloom's syndrome helicase; recQ helicase homology
C/Keywords: ATP, nucleic acid binding, P-loop
F/541-548/Region: nucleotide-binding motif A (P-loop)
F/647-652/Region: nucleotide-binding motif B
F/651-654/Region: DEAH motif
F/875-914/Domain: recQ helicase homology <RHH>

Query Match      8.2%; Score 525.5; DB 2; Length 1328;
Best Local Similarity 37.1%; Pred. No. 2.8e-20;
Matches 143; Conservative 49; Mismatches 154; Indels 39; Gaps 11;

QY      443 VPSLDPYVLYPLYSIGSPGQLAETPAEYVQALQEGH--QAFPRGGERANMRLTSGISTILV 501
      :| | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

Search completed: December 27, 2005, 21:54:05
Job time : 48.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: December 27, 2005, 21:38:41; Search time 216.8 seconds

(without alignments)
2448.199 Million cell updates/sec

Title: US-09-889-325-4

Perfect score: 6424

Sequence: 1 MERLRDVRERQAMERAFRR.....HLSTHALVGLATEELLQVAR 1208

Scoring table: BLOSUM62

GAPOP 10.0, Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: Geneseq.21:*
2: geneseqp1980s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003as:*
8: geneseqp2003bs:*
9: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6424	100.0	1208	3	AA020993 Human Rec
2	6424	100.0	1208	5	AB033387 Human Rec
3	6424	100.0	1208	8	ADQ21562 Human Rec
4	6424	100.0	1208	8	ADU80701 Human Rec
5	6421	100.0	1208	2	AAW95050 Human hel
6	6085.5	94.7	1306	8	ADM90956 Human pha
7	6050	94.2	1144	7	ADD01202 Human nuc
8	6029	93.9	1142	8	ABM84821 Human dia
9	4006	62.4	756	3	AA020995 Human Rec
10	3944	61.4	1216	9	AAW44021 Mouse REC
11	2920	45.5	554	3	AA020994 Human Rec
12	1723.5	26.8	1579	4	AB063317 Drosophila
13	631.5	9.8	610	4	AAU34821 E. coli c
14	631.5	9.8	610	6	ABU15433 Protein e
15	619	9.6	609	6	ABU28286 Protein e
16	609.5	9.5	609	6	ABU47859 Protein e
17	609.5	9.5	615	4	AAU38154 Salmonella
18	607.5	9.5	615	6	ABU47049 Protein e
19	606	9.4	608	7	ABU23005 Protein e
20	606	9.4	619	7	AB061558 Kiebsiella
21	593	9.2	610	6	ABU50101 Protein e
22	591.5	9.2	641	7	ADP04074 Bacterial
23	589.5	9.2	600	6	ABU41290 Protein e
24	586.5	9.1	613	6	ABM68051 Phototaph

25	585	9.1	1394	9	ABM97298 M. xanthu
26	581.5	9.1	620	6	ABU49054 Protein e
27	578.5	9.0	632	6	ABU39369 Protein e
28	578	9.0	619	4	AAU35492 Haemophil
29	578	9.0	619	6	ABU10347 Protein e
30	554.5	8.6	589	6	ABU14523 Protein e
31	554.5	8.6	589	7	ADH87825 Enterococ
32	553	8.6	608	9	AB042257 L. pneumo
33	553	8.6	608	9	AB039122 L. pneumo
34	552	8.6	590	5	AB048903 Listeria
35	552	8.6	590	6	ABU32767 Protein e
36	549	8.5	597	6	ABU44940 Protein e
37	547.5	8.5	589	4	AAU35210 Enterococ
38	547	8.5	849	7	ABO69207 Pseudomon
39	545	8.5	580	4	AAU33463 Enterococ
40	545	8.5	712	4	AAU36322 Pseudomon
41	545	8.5	712	6	ABU38552 Protein e
42	544.5	8.5	390	6	ABU33517 Protein e
43	544	8.5	341	8	ADN26337 Bacterial
44	542.5	8.4	1955	3	AAU94670 Silencing
45	536	8.3	651	5	ABP65937 Bifidobac

ALIGNMENTS

RESULT 1
AAB20993
ID AAB20993 standard; protein, 1208 AA.
XX
AC AAB20993;
XX
DT 11-DEC-2000 (first entry)
XX
DE Human RecQ4 helicase.
XX
KW RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;
KW polkioderma congenitale; autosomal recessive; skin disorder;
KW dermatology; antibody; prenatal diagnosis; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200043522-A1.
XX
PD 27-JUL-2000.
XX
PF 19-JAN-2000; 2000WO-JP000233.
XX
PR 19-JAN-1999; 99JP-00011218.
XX
PA (AGEN-) AGENE RES INST CO LTD.
XX
PI Kitao S, Shimamoto A, Furuchi Y;
XX
DR MPI; 2000-524241/47.
XX
PT N-PSDB; AAA72320, AAA72321.
XX
PS Example 5; Page 83-92; 115pp; Japanese.
XX
CC The present sequence represents human RecQ4 helicase. The invention
CC relates to the genomic DNA sequence of human RecQ4 helicase (AAA72320).
CC Mutations in the RecQ4 helicase gene, located on chromosome 8q24.3, are
CC the cause of Rothmund-Thomson syndrome (also known as poikiloderma
CC congenitale), an autosomal recessive skin disorder principally occurring
CC in females and often accompanied by juvenile cataracts, saddle nose,
CC congenital bone defects, hypogonadism and disturbances in the growth of
CC hair, nails and teeth. The invention also relates to vectors and host
CC cells comprising the human RecQ4 helicase genomic sequence. It
CC additionally encompasses use of the RecQ4 helicase protein as a
CC therapeutic and anti-Reco4 antibodies as diagnostic agents. The RecQ4

CC helicase gene and its products, and anti-RecoQ4 helicase antibodies are
CC useful in the diagnosis, especially prenatal diagnosis, and treatment of
CC Rothmund-Thomson syndrome. The genomic sequence may especially be used in
CC gene therapy for this condition

XX Sequence 1208 AA;

Query Match 100.0%; Score 6424; DB 3; Length 1208;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MERLDVNERLOAERARQRGRPPSODDVEAAPEETRALYREYRTLKRTTGAAGGLR 60
DB 1 MERLDVNERLOAERARQRGRPPSODDVEAAPEETRALYREYRTLKRTTGAAGGLR 60
QY 61 SSESIPAAAEAPERPCCWPHLNRAATYSPOPTGERSQGSVPDYQGLKANLKTGLAQG 120
DB 61 SSESIPAAAEAPERPCCWPHLNRAATYSPOPTGERSQGSVPDYQGLKANLKTGLAQG 120
QY 121 PALGRRPWPLGRASSKASTPKPGTGPVPSFAKXVSDSPOLPEOPRGRLOHLQASLS 180
DB 121 PALGRRPWPLGRASSKASTPKPGTGPVPSFAKXVSDSPOLPEOPRGRLOHLQASLS 180
QY 181 QRLGSLDPGWLQRCHEVDFLGAPKACRPDLGSEESQLLPGESAVLGPAGSGQPEAS 240
DB 181 QRLGSLDPGWLQRCHEVDFLGAPKACRPDLGSEESQLLPGESAVLGPAGSGQPEAS 240
QY 241 AFQSVSTRVSGPSSSGGGRKRWNEBWPESPAVOQSSSQAGPSEAGAVAEEDPG 300
DB 241 AFQSVSTRVSGPSSSGGGRKRWNEBWPESPAVOQSSSQAGPSEAGAVAEEDPG 300
QY 301 BPVQAQPPQPCSSPSNPRHYGLSPSSQARAGAEGLTAFLHFPRLARHDRGNVYRLNMKQ 360
DB 301 BPVQAQPPQPCSSPSNPRHYGLSPSSQARAGAEGLTAFLHFPRLARHDRGNVYRLNMKQ 360
QY 361 KHYVGRALRSRLLRKQAKQKMRKQECFCGGGATVTTKESCFLEQFDHMAAQCPRA 420
DB 361 KHYVGRALRSRLLRKQAKQKMRKQECFCGGGATVTTKESCFLEQFDHMAAQCPRA 420
QY 421 SEEDTDVNGEPPLVPSPPVPEVPSLDPTVPLYSLSGSGQALSTPAVFOALRQLGHOA 480
DB 421 SEEDTDVNGEPPLVPSPPVPEVPSLDPTVPLYSLSGSGQALSTPAVFOALRQLGHOA 480
QY 481 FRPQGERAVMRLISGISITLVLPAGAKSLCYOLPALIYSRSPCLTVSPILSIMDQ 540
DB 481 FRPQGERAVMRLISGISITLVLPAGAKSLCYOLPALIYSRSPCLTVSPILSIMDQ 540
QY 541 VSGLPCLKAKACIHSGMTKQRESVLQKIRAAQVHVLMTPEALVGAAGLPAAQLPPVA 600
DB 541 VSGLPCLKAKACIHSGMTKQRESVLQKIRAAQVHVLMTPEALVGAAGLPAAQLPPVA 600
QY 601 FACIDEAHCLSQSHNPRPCTLRVCVKTLREMGVHCFGLTATATRTTASVVAQHLAAE 660
DB 601 FACIDEAHCLSQSHNPRPCTLRVCVKTLREMGVHCFGLTATATRTTASVVAQHLAAE 660
QY 661 EPDLHGPAVPVNTNHLSTVSMRDTDOALLTLQGRFQNDLSIIYYCRRRDETERIALL 720
DB 661 EPDLHGPAVPVNTNHLSTVSMRDTDOALLTLQGRFQNDLSIIYYCRRRDETERIALL 720
QY 721 RTCLHAAMVPSGGARPKTTAEAYHAGCSERRRRVQAFMGQQLRVVATVAFMGCLDR 780
DB 721 RTCLHAAMVPSGGARPKTTAEAYHAGCSERRRRVQAFMGQQLRVVATVAFMGCLDR 780
QY 781 PDVAVALHGLPSPFESTVQAVGARGDQPAHCHLFTQPOGEDLRELRHHVADSTDFL 840
DB 781 PDVAVALHGLPSPFESTVQAVGARGDQPAHCHLFTQPOGEDLRELRHHVADSTDFL 840
QY 841 AVKLVORVPACCTCTRPSEBEGANVGERPVKYPPOAEOLSHQAAAGPRVCMGH 900
DB 841 AVKLVORVPACCTCTRPSEBEGANVGERPVKYPPOAEOLSHQAAAGPRVCMGH 900
QY 901 ERALPIQLTVQALDMPBEALETLLCYLELPHHMLLELATYTHCRNLNCPGPAQLQALA 960
DB 901 ERALPIQLTVQALDMPBEALETLLCYLELPHHMLLELATYTHCRNLNCPGPAQLQALA 960
```

```
DB 901 ERALPIQLTVQALDMPBEALETLLCYLELPHHMLLELATYTHCRNLNCPGPAQLQALA 960
QY 961 HRCPPILAVCLAQQLPEDPGQSSSVFEDMYKLVDSKMEILASVRRALCOLQMDHEPRTGV 1020
DB 961 HRCPPILAVCLAQQLPEDPGQSSSVFEDMYKLVDSKMEILASVRRALCOLQMDHEPRTGV 1020
QY 1021 RRGTVLVPSFELAFHRSRPGDLTAERKDOI CDPLVGRVQAREQALARLRTRTQAHSV 1080
DB 1021 RRGTVLVPSFELAFHRSRPGDLTAERKDOI CDPLVGRVQAREQALARLRTRTQAHSV 1080
QY 1081 AFPSCGPLEQODEERSTRKDLGRFYEEBEGQEPGAMEDAQPPEGOARLODMEQVR 1140
DB 1081 AFPSCGPLEQODEERSTRKDLGRFYEEBEGQEPGAMEDAQPPEGOARLODMEQVR 1140
QY 1141 CDINQFSLRPEEKFSRAVARIFHGIGPCYPAQVYGODRRFMRKYLHLSFHALVGLAT 1200
DB 1141 CDINQFSLRPEEKFSRAVARIFHGIGSPCYPAQVYGODRRFMRKYLHLSFHALVGLAT 1200
QY 1201 BELLOVAR 1208
DB 1201 BELLOVAR 1208

RESULT 2
ABG93387
ID ABG93387 standard; protein; 1208 AA.
XX
AC ABG93387;
XX
DT 20-NOV-2002 (first entry)
XX
DE Human RecoQ protein-like 4 (RECOQ4).
XX
KW Human; RecoQ protein-like 4; RECOQ4; infection; inflammation; cytostatic;
KW tumour formation; cancer; antiinflammatory; antimicrobial;
KW antisense therapy.
XX
OS Homo sapiens.
XX
PN US6436706-B1.
XX
PD 20-AUG-2002.
XX
PF 23-FEB-2001; 2001US-00792594.
XX
PR 23-FEB-2001; 2001US-00792594.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Ward DT, Watt AT;
XX
DR WPI; 2002-689941/74.
DR N-PSDB; ABS68851.
XX
PT New antisense compounds targeted to nucleic acid encoding RecoQ protein-
PT like 4, useful for modulating expression of the nucleic acid and treating
PT diseases associated with expression of the nucleic acid in humans.
XX
PS Disclosure; Col 47-56; 45pp; English.
XX
CC The invention relates to a compound targeted to specific nucleobases of
CC RecoQ protein-like 4 (RECOQ4) and which hybridises and inhibits the
CC expression of RECOQ4. The compound is useful for inhibiting the
CC expression of RECOQ4 in cells or tissues and for treating an animal,
CC particularly a human suspected of having or being prone to a disease or
CC condition associated with expression of RECOQ4. The compound is useful
CC for diagnostics, therapeutics and as a research reagent, e.g.
CC prophylactically to prevent or delay infection, inflammation or tumour
CC formation. This sequence represents human RECOQ4 polypeptide
CC
XX Sequence 1208 AA;
```

Query Match 100.0%; Score 6424; DB 5; Length 1208;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MERLDVRRERLOAMERAFRRORGRPSQDDVAABETRALYREYRTTKRTTGOAGGGR 60
DB 1 MERLDVRRERLOAMERAFRRORGRPSQDDVAABETRALYREYRTTKRTTGOAGGGR 60
QY 61 SEESLPAABEABERPCWGPBHLNRATKSPPTPGSRGQSVDPGORKANLKGTLQAG 120
DB 61 SEESLPAABEABERPCWGPBHLNRATKSPPTPGSRGQSVDPGORKANLKGTLQAG 120
QY 121 PALGRRPWPPLGRASSKASTPKPPGTGVPVFAKTVSDPEPOLPEPPRPRRLQHLQASIS 180
DB 121 PALGRRPWPPLGRASSKASTPKPPGTGVPVFAKTVSDPEPOLPEPPRPRRLQHLQASIS 180
QY 181 QRLGSLDPGWLORCHSEVPDPLGAPKACRPDLGSEBSQLLIPGESAVLGPAGASQGPBAS 240
DB 181 QRLGSLDPGWLORCHSEVPDPLGAPKACRPDLGSEBSQLLIPGESAVLGPAGASQGPBAS 240
QY 241 AROEVSIRVGSPOSSSGGKKRMNEEPWESPAOVQOESSQAGPPEEGAGAVAEEDPPG 300
DB 241 AROEVSIRVGSPOSSSGGKKRMNEEPWESPAOVQOESSQAGPPEEGAGAVAEEDPPG 300
QY 301 BEVOAOPPOPCSSPNRYHGLSPSSQABAGKAGTAPLHIFPRLARHDRGNVYRLNMKQ 360
DB 301 BEVOAOPPOPCSSPNRYHGLSPSSQABAGKAGTAPLHIFPRLARHDRGNVYRLNMKQ 360
QY 361 KHYVRGRALRSRLRKQANKQKRRKKGECFGGGAGATVTTKESCFLEQPDHMAAOCPRA 420
DB 361 KHYVRGRALRSRLRKQANKQKRRKKGECFGGGAGATVTTKESCFLEQPDHMAAOCPRA 420
QY 421 SEEDTDVAVPEPLVPEPQVPEPVLDPVLYPLYSIGPSGQLAETAEVFALEEQGHQA 480
DB 421 SEEDTDVAVPEPLVPEPQVPEPVLDPVLYPLYSIGPSGQLAETAEVFALEEQGHQA 480
QY 481 PRPGQERAWRILISGSTLLVLTPTGAGKSLCYOLPALIYSRSPCLTLVSPPLSLMDQ 540
DB 481 PRPGQERAWRILISGSTLLVLTPTGAGKSLCYOLPALIYSRSPCLTLVSPPLSLMDQ 540
QY 541 VSGLPPECLAKACIHSQMTKQRESVLQKIRAAQVHVLMTPEALVAGGIPPAQOLPVA 600
DB 541 VSGLPPECLAKACIHSQMTKQRESVLQKIRAAQVHVLMTPEALVAGGIPPAQOLPVA 600
QY 601 FACIDBAHCLSQSNHPRPCYLAVCVLAERBQVHCFGLTATARTASDVQAHLAAVE 660
DB 601 FACIDBAHCLSQSNHPRPCYLAVCVLAERBQVHCFGLTATARTASDVQAHLAAVE 660
QY 661 EBDLHGPAPVPTLHLSVSMRDTDOALLTLQKSKFQNLDSIIICNRREDTERIALLL 720
DB 661 EBDLHGPAPVPTLHLSVSMRDTDOALLTLQKSKFQNLDSIIICNRREDTERIALLL 720
QY 721 RTCLHAAMVSGSGRAPKTTAAVYHAGMCSRRERRVQRAFQOLNVVATVAFGKGLDR 780
DB 721 RTCLHAAMVSGSGRAPKTTAAVYHAGMCSRRERRVQRAFQOLNVVATVAFGKGLDR 780
QY 781 PDVRAVLAHLGLPSFESSYQAVARAGDQPAHCHFLQOGSEDLRLRHVHADSTDL 840
DB 781 PDVRAVLAHLGLPSFESSYQAVARAGDQPAHCHFLQOGSEDLRLRHVHADSTDL 840
QY 841 AVKRLVORVPACTCTPRPSQEGAVGERVVPKPPPEAQLSHQAAPGRVCMGH 900
DB 841 AVKRLVORVPACTCTPRPSQEGAVGERVVPKPPPEAQLSHQAAPGRVCMGH 900
QY 901 ERALPTQLTVALDMBEAEITLLCYELHPHMLBELATTYHCRINCPCGPAQIQALA 960
DB 901 ERALPTQLTVALDMBEAEITLLCYELHPHMLBELATTYHCRINCPCGPAQIQALA 960
QY 961 HRCPEPLAVCLAQOLPEDPGQSSVEFDWYKLVDSQMGWELASVRBALCOLQMDHBPRTGV 1020
DB 961 HRCPEPLAVCLAQOLPEDPGQSSVEFDWYKLVDSQMGWELASVRBALCOLQMDHBPRTGV 1020
QY 1021 RRGTVLVEFSELAFLHLSRPGDLTAEEKQICDPLVGRVQARERQALRLRTTQAFHSV 1080
DB 1021 RRGTVLVEFSELAFLHLSRPGDLTAEEKQICDPLVGRVQARERQALRLRTTQAFHSV 1080

```

Db 1021 RRGTVLVEFSELAFLHLSRPGDLTAEEKQICDPLVGRVQARERQALRLRTTQAFHSV 1080
QY 1081 AFGSGPCLTEQODERSTRLLDLGRYFEEEGGEPGGMEDAOGPEPQALQDMEQVR 1140
Db 1081 AFGSGPCLTEQODERSTRLLDLGRYFEEEGGEPGGMEDAOGPEPQALQDMEQVR 1140
QY 1141 CDIRQFLSLRPEEKSSRAVARIFHGIGSPCYPAQVYGODRRFWRKYLHLSFHALVGLAT 1200
Db 1141 CDIRQFLSLRPEEKSSRAVARIFHGIGSPCYPAQVYGODRRFWRKYLHLSFHALVGLAT 1200
QY 1201 EBLQVAR 1208
Db 1201 EBLQVAR 1208

RESULT 3
ADQ21562
ID ADQ21562 standard; protein; 1208 AA.
XX
AC ADQ21562;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 4382.
XX
KW soft tissue sarcoma; cytoskeletal; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PE 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX

PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 4382; 210pp; English.
XX

CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytoskeletal applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX

Sequence 1208 AA;
SQ

Query Match 100.0%; Score 6424; DB 8; Length 1208;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MERLDVRRERLOAMERAFRRORGRPSQDDVAABETRALYREYRTTKRTTGOAGGGR 60
DB 1 MERLDVRRERLOAMERAFRRORGRPSQDDVAABETRALYREYRTTKRTTGOAGGGR 60

```



```

QY 1 MERLADVVERLOAMERAFRRORGRSPSODDVEAEAEETRALYREYRTLKRTTGOAGGJR 60
DB 1 MERLADVVERLOAMERAFRRORGRSPSODDVEAEAEETRALYREYRTLKRTTGOAGGJR 60
QY 61 SSESIPAAAEAEAEPRCWPGLNRAATKSPQTPGRSRQGSVPDYGOGLKANLKGTLQAG 120
DB 61 SSESIPAAAEAEAEPRCWPGLNRAATKSPQTPGRSRQGSVPDYGOGLKANLKGTLQAG 120
QY 121 PALGRPMPPLGASAKASTPRPGTGPVPSFAEKYSDEBPOLPEPQPRGRLOHLQASLS 180
DB 121 PALGRPMPPLGASAKASTPRPGTGPVPSFAEKYSDEBPOLPEPQPRGRLOHLQASLS 180
QY 181 ORLGSLDGMLORCSEVPDFLGAPKACRPDLGSEESQLITPGESAVLGPAGASQGPAS 240
DB 181 ORLGSLDGMLORCSEVPDFLGAPKACRPDLGSEESQLITPGESAVLGPAGASQGPAS 240
QY 241 AFQEVSIIVGSPQSSSGGKKRMBEPWESPAVOVQOSSQAGPPSEGAGAVAEDEPPG 300
DB 241 AFQEVSIIVGSPQSSSGGKKRMBEPWESPAVOVQOSSQAGPPSEGAGAVAEDEPPG 300
QY 301 BPVQAQPPQPCSSBNPRYHGISSESSQARAGAEGLAHLTPRLARHDKGNYVRLNMKQ 360
DB 301 BPVQAQPPQPCSSBNPRYHGISSESSQARAGAEGLAHLTPRLARHDKGNYVRLNMKQ 360
QY 361 KHYVGRALRSRLRKQAMKOKMKRKKGCGFGGGAATVTTKSCFLNEOPDHMAAOCPPRA 420
DB 361 KHYVGRALRSRLRKQAMKOKMKRKKGCGFGGGAATVTTKSCFLNEOPDHMAAOCPPRA 420
QY 421 SEEDTDVAGPEPLVPSPOVPEVPSLDPTVPLVSLSGSGQLAETPAEYFOALBQLGHA 480
DB 421 SEEDTDVAGPEPLVPSPOVPEVPSLDPTVPLVSLSGSGQLAETPAEYFOALBQLGHA 480
QY 481 FRPGQERAVMRLTSGISTLLVPTGAGKSLCYQLPALYSRSRSCITLVSVLSLMDQ 540
DB 481 FRPGQERAVMRLTSGISTLLVPTGAGKSLCYQLPALYSRSRSCITLVSVLSLMDQ 540
QY 541 VSGLPCLCAACIHSGMTRKORESVLOKIRAAOVHMLTPRALVAGAGGLPPAAOLPPVA 600
DB 541 VSGLPCLCAACIHSGMTRKORESVLOKIRAAOVHMLTPRALVAGAGGLPPAAOLPPVA 600
QY 601 FACIDEAHCLSQMSHNFPCYLKVCVLEEMGVHCFGLTATARTASDVAAOHLAVAE 660
DB 601 FACIDEAHCLSQMSHNFPCYLKVCVLEEMGVHCFGLTATARTASDVAAOHLAVAE 660
QY 661 EPDLHGPAVPPTNLHLSVSMRDTQALLTLQGRFQNLDSIIYCNREDETERIALLL 720
DB 661 EPDLHGPAVPPTNLHLSVSMRDTQALLTLQGRFQNLDSIIYCNREDETERIALLL 720
QY 721 RFLCLAAAWPGSGGAPKTTAAVYHAGMSRRRRVQRAFMQGLAVVVAATVAFQMGILDR 780
DB 721 RFLCLAAAWPGSGGAPKTTAAVYHAGMSRRRRVQRAFMQGLAVVVAATVAFQMGILDR 780
QY 781 PVRRAVTHLGLPSPFESYVQAVGRAGRDQPAHCHFLPOGSEDLRELRHHVHADSTDL 840
DB 781 PVRRAVTHLGLPSPFESYVQAVGRAGRDQPAHCHFLPOGSEDLRELRHHVHADSTDL 840
QY 841 AYKRLVORVPACTCTCTPRPSBOGAVGERPVKYPQOEAELSHOAPGRRVCMGH 900
DB 841 AYKRLVORVPACTCTCTPRPSBOGAVGERPVKYPQOEAELSHOAPGRRVCMGH 900
QY 901 ERALPQLVOLADMEEALETLLCYLEHPPHMLLATTYTHGCLNCGGPAOLQALA 960
DB 901 ERALPQLVOLADMEEALETLLCYLEHPPHMLLATTYTHGCLNCGGPAOLQALA 960
QY 961 HRCPLAVCLAOQLPEDPGQSSSVFEDWVKLVDSNGWELASVRALCOLQMDHEBRTGV 1020
DB 961 HRCPLAVCLAOQLPEDPGQSSSVFEDWVKLVDSNGWELASVRALCOLQMDHEBRTGV 1020
QY 1021 RRGCTGLVPSRLAFLBSRGDLTAEKQICDPLVGRVQARRQALALRRFQAFHSV 1080
DB 1021 RRGCTGLVPSRLAFLBSRGDLTAEKQICDPLVGRVQARRQALALRRFQAFHSV 1080

```

```

QY 1081 AFPSCGPLEQODEERSTRKDLIGRYFEEBEGEPGAMEDAQCPBGQARLODMEDQVR 1140
DB 1081 AFPSCGPLEQODEERSTRKDLIGRYFEEBEGEPGAMEDAQCPBGQARLODMEDQVR 1140
QY 1141 CDIQFSLRPEEFSSRAVARIFHIGISPCYPAOVYGOBRFMRKYLHLSEHLLVGLAT 1200
DB 1141 CDIQFSLRPEEFSSRAVARIFHIGISPCYPAOVYGOBRFMRKYLHLSEHLLVGLAT 1200
QY 1201 BELLOVAR 1208
DB 1201 BELLOVAR 1208

RESULT 5
ID AAM95050
AAM95050 standard; protein; 1208 AA.
XX
AC AAM95050;
XX
DT 14-MAY-1999 (first entry)
XX
DE Human helicase protein.
XX
KW RecQ4 gene; helicase; Werner's syndrome; Bloom's syndrome; human.
XX
OS Homo sapiens.
XX
PN W0905284-A1.
XX
PD 04-FEB-1999.
XX
PF 10-JUL-1998; 98WO-JP003114.
XX
PR 25-JUL-1997; 97JP-00200387.
XX
PA (AGEN-) AGENE RES INST CO LTD.
XX
PI Shimamoto A, Kitao S, Furuichi Y;
XX
DR WPI; 1999-142939/12.
XX
DR N-PSDB; AAX21656.
XX
PT New human helicase gene RecQ4 - used for investigation and diagnosis of
PS helicase-implicated diseases such as Werner's syndrome.
XX
PS Claim 1; Page 35-42; 67pp; Japanese.
XX
CC The present sequence represents a protein having helicase activity
CC encoded by the human gene RecQ4. The gene has significant homology to the
CC Escherichia coli helicase gene (RecQ). Host cells transformed with
CC vectors comprising the RecQ4 gene are used for the recombinant expression
CC of the protein. The gene may be used for the study and diagnosis of
CC disorders in which helicase activity is involved, such as Werner's and
CC Bloom's syndromes in which mutations in the helicase gene are implicated
XX
SQ Sequence 1208 AA;

Query Match 100.0%; Score 6421; DB 2; Length 1208;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1207; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MERLADVVERLOAMERAFRRORGRSPSODDVEAEAEETRALYREYRTLKRTTGOAGGJR 60
DB 1 MERLADVVERLOAMERAFRRORGRSPSODDVEAEAEETRALYREYRTLKRTTGOAGGJR 60
QY 61 SSESIPAAAEAEAEPRCWPGLNRAATKSPQTPGRSRQGSVPDYGOGLKANLKGTLQAG 120
DB 61 SSESIPAAAEAEAEPRCWPGLNRAATKSPQTPGRSRQGSVPDYGOGLKANLKGTLQAG 120
QY 1201 PALGRPMPPLGASAKASTPRPGTGPVPSFAEKYSDEBPOLPEPQPRGRLOHLQASLS 180
DB 1201 PALGRPMPPLGASAKASTPRPGTGPVPSFAEKYSDEBPOLPEPQPRGRLOHLQASLS 180

```

181 ORLSLDPGWLORCHSEVDFLAGPACRPRDUGSEESOLLIRGSANVLGPAGSGCPRAS 240
181 ORLSLDPGWLORCHSEVDFLAGPACRPRDUGSEESOLLIRGSANVLGPAGSGCPRAS 240
241 AFOEVSIRVSGPOSSSGEKRRNMEEPMESPAQVOQESSQAGPSEBAGAVAVEDEPPG 300
241 AFOEVSIRVSGPOSSSGEKRRNMEEPMESPAQVOQESSQAGPSEBAGAVAVEDEPPG 300
301 EPVQAQPPQPCSSSPNPRYHGLSPSSQARAKABETALHIFPRLARHDRKNTYVTLNKKQ 360
301 EPVQAQPPQPCSSSPNPRYHGLSPSSQARAKABETALHIFPRLARHDRKNTYVTLNKKQ 360
361 KHYRGRALRERLLRKQAKWKRRKGCFCGCGGATVTTKISCFINQEPDHAACCPSPA 420
361 KHYRGRALRERLLRKQAKWKRRKGCFCGCGGATVTTKISCFINQEPDHAACCPSPA 420
421 SEEDTDVAGPEPLVPSPOPEVEPSLDPTVLPLVSLGSGQLAETPAEVFOALBQLGHQA 480
421 SEEDTDVAGPEPLVPSPOPEVEPSLDPTVLPLVSLGSGQLAETPAEVFOALBQLGHQA 480
481 FRPQGERAVNRILSGISTLVLPTGAGKSLCYQLPALIYSRRSPCLTLVSPPLSLMDQ 540
481 FRPQGERAVNRILSGISTLVLPTGAGKSLCYQLPALIYSRRSPCLTLVSPPLSLMDQ 540
541 VSGPPLCLKAACINSGMTRKQRESVLQKIRAAQVHVLMTPEALVAGAGLPPAQLPPVA 600
541 VSGPPLCLKAACINSGMTRKQRESVLQKIRAAQVHVLMTPEALVAGAGLPPAQLPPVA 600
601 PACIDEAHCLSGWSHNFRCYLAVCKVLBERGVHCFGLTATATRTASDVACHLAVAE 660
601 PACIDEAHCLSGWSHNFRCYLAVCKVLBERGVHCFGLTATATRTASDVACHLAVAE 660
661 EPDLHGPAVPVTNHLISVMDRDTQALLTLQGRFONDLSIIYCNRRBETTERIALLL 720
661 EPDLHGPAVPVTNHLISVMDRDTQALLTLQGRFONDLSIIYCNRRBETTERIALLL 720
721 RTCLHAAPVPSGGGRAPKTTABAYHAGCSRRERRVQAFQOGOLRVVAVTAVFEMGDR 780
721 RTCLHAAPVPSGGGRAPKTTABAYHAGCSRRERRVQAFQOGOLRVVAVTAVFEMGDR 780
781 PDVRAVHLGLPSPESVYQAVGRAGRDQPAHCHLFLQPGQEDRELRRHVHADSTDFL 840
781 PDVRAVHLGLPSPESVYQAVGRAGRDQPAHCHLFLQPGQEDRELRRHVHADSTDFL 840
841 AVKRLVQRVFPACTCTCTRPSEBQAGVGERPVKTPPOABEQLSHQAAFGPRVCMGH 900
841 AVKRLVQRVFPACTCTCTRPSEBQAGVGERPVKTPPOABEQLSHQAAFGPRVCMGH 900
901 ERALPIQLTVQALDMPBEAIEFTLLCYELHHPHMEELATTYTHCRNLCPGSPQOLQALA 960
901 ERALPIQLTVQALDMPBEAIEFTLLCYELHHPHMEELATTYTHCRNLCPGSPQOLQALA 960
961 HRCPEPLAVCLAQOLPEDPGQSSSVFDMVKLVDSMGWELASVRRALQOLQMDHEPRIGV 1020
961 HRCPEPLAVCLAQOLPEDPGQSSSVFDMVKLVDSMGWELASVRRALQOLQMDHEPRIGV 1020
1021 RRGTVGVVESELAFLHLSPGDLTAEEKQICDPLFYGVQABERQALARLRTQAFHSV 1080
1021 RRGTVGVVESELAFLHLSPGDLTAEEKQICDPLFYGVQABERQALARLRTQAFHSV 1080
1081 AFPSGCPCLGQODERSTRKDLGRVFEESGQPGMEQAGPEQOALQWEDQV 1140
1081 AFPSGCPCLGQODERSTRKDLGRVFEESGQPGMEQAGPEQOALQWEDQV 1140
1141 CDIROPFLSRBEKFSRAVARIFHGIGSPCYAQVYQODRRFMRKYLHLSFHALVGLAT 1200
1141 CDIROPFLSRBEKFSRAVARIFHGIGSPCYAQVYQODRRFMRKYLHLSFHALVGLAT 1200
1201 BELIQVAR 1208
1201 BELIQVAR 1208
1201 BELIQVAR 1208
1201 BELIQVAR 1208

RESULT 6
ADM90956
ID ADM90956 standard; protein; 1306 AA.
XX
XX
AC ADM90956;
XX
XX
DT 03-JUN-2004 (first entry)
XX
DE Human pharmaceutically useful protein SeqID 349.
XX
XX human; cancer; haematopoiesis; thrombosis; anaemia;
KW cardiovascular disorder; ischemic heart disease;
KW acute myocardial infarction; respiratory disease; asthma; pneumonia;
KW cyclic fibrosis; chronic renal failure; glomerulopathy;
KW gastrointestinal disorder; peptic ulcer; cirrhosis; immune disorder;
KW HIV infection; systemic lupus erythematosus; endocrine system;
KW diabetes mellitus; epilepsy; Alzheimer's disease;
KW amyotrophic lateral sclerosis; skin disorder; psoriasis; bacterial;
KW fungal; parasitic; viral infection; cytostatic; anticoagulant;
KW thrombolytic; antianaemic; cardiant; vasostrictic; antiasthmatic;
KW antiinflammatory; nephrotropic; antiulcer; hepatotropic;
KW immunosuppressive; antiallergic; dermatological; antirheumatic;
KW antiarthritic; antidiabetic; anticonvulsant; neuroprotective; nootropic;
KW antipariatic; antibacterial; fungicide; antiparasitic; virucidal;
KW gene therapy; vaccine.
XX
XX Homo sapiens.
XX
XX WO2004020595-A2.
XX
XX 11-MAR-2004.
XX
XX 28-AUG-2003; 2003WO-US027107.
XX
XX 29-AUG-2002; 2002US-0406576P.
XX 29-AUG-2002; 2002US-0406611P.
XX 29-AUG-2002; 2002US-0406612P.
XX 29-AUG-2002; 2002US-0406616P.
XX 29-AUG-2002; 2002US-0406640P.
XX 29-AUG-2002; 2002US-0406655P.
XX 29-AUG-2002; 2002US-0406666P.
XX 17-SEP-2002; 2002US-0410946P.
XX 17-SEP-2002; 2002US-0410951P.
XX 17-SEP-2002; 2002US-0410953P.
XX 17-SEP-2002; 2002US-0410957P.
XX 17-SEP-2002; 2002US-0410962P.
XX 17-SEP-2002; 2002US-0410962P.
XX 17-SEP-2002; 2002US-0411019P.
XX 17-SEP-2002; 2002US-0411022P.
XX 17-SEP-2002; 2002US-0411024P.
XX 17-SEP-2002; 2002US-0411037P.
XX 17-SEP-2002; 2002US-0411046P.
XX 17-SEP-2002; 2002US-0411052P.
XX 17-SEP-2002; 2002US-0411082P.
XX 17-SEP-2002; 2002US-0411111P.
XX
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.
XX (RIKE-) RIKEN INST PHYSICAL & CHEM RES.
XX (DNAF-) DNAFORM KK.
XX
XX Williams LT, Chu K, Lee E, Heecir K;
XX
XX WPI; 2004-257410/24.
XX N-PSDB; ADM90747, ADM91165.
XX
XX New human polynucleotides and polypeptides, useful for diagnosing,
XX preventing and treating proliferative disorders, immune disorders,
XX cardiovascular disorders, or bacterial, fungal, parasitic and viral
XX diseases.
XX
XX Claim 1; SEQ ID NO 349; 254pp; English.
XX
XX This invention relates to novel isolated human polynucleotides and the

CC encoded proteins thereof. Specifically, it refers to proteases, kinases, CC phosphatases, secreted and transmembrane proteins, as well as the derived CC peptide fragments, which can be used to develop antibodies and screen for CC small molecule agonists and antagonists that can modulate their CC activities. The present invention describes polypeptides, CC polynucleotides, vectors and host cells useful for diagnosing, preventing CC and treating proliferative disorders, e.g. cancer, disorders of CC haematopoiesis such as thrombosis and anaemia, cardiovascular disorders, CC e.g. ischaemic heart disease and acute myocardial infarction, respiratory diseases, e.g. asthma, pneumonia or cystic fibrosis, disorders of the CC kidney and urinary tract, e.g. chronic renal failure and glomerulopathy, CC gastrointestinal disorders, e.g. peptic ulcer or cirrhosis, immune CC disorders, e.g. HIV infection and systemic lupus erythematosus, disorders of the endocrine system, e.g. diabetes mellitus, central nervous system disorders, e.g. epilepsy, Alzheimer's disease or amyotrophic lateral sclerosis, skin disorders, e.g. psoriasis, as well as bacterial, fungal, CC parasitic and viral diseases. Accordingly, they exhibit many various CC activities including cytostatic, anticoagulant, thrombolytic, CC antinaeemic, cardiant, vasotropic, antiaesthetic, antiinflammatory, CC nephrotropic, anticancer, hepatotropic, immunosuppressive, antifungal, CC dermatological, antineumatic, antiarthritic, antidiabetic, antiallergic, CC anticonvulsant, neuroprotective, nootropic, antipsoriatic, antibacterial, CC fungicide, antiparasitic and virucidal, such that these polynucleotides CC can be used for gene therapy purposes and the development of appropriate CC vaccines. This polypeptide is a human protein of the invention.

XX Sequence 1306 AA;

Query Match 94.7%; Score 6085.5; DB 8; Length 1306;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1167; Conservative 2; Mismatches 26; Indels 65; Gaps 6;

QY 1 MERLNVRERLQWAEAFRRGRGRSPQDVAAEETALREYTLKRTTQAGAGGR 60
DB 60 MERLNVRERLQWAEAFRRGRGRSPQDVAAEETALREYTLKRTTQAGAGGR 119
QY 61 SSESPLAAEAEPEPCWPHLNRAATKSPQPTPGSRQGVDPYQRLKAMIKGTLQAG 120
DB 120 SSESPLAAEAEPEPCWPHLNRAATKSPQPTPGSRQGVDPYQRLKAMIKGTLQAG 179
QY 121 PALGRAPWPLGRASSASTPKEPGTGVPSFAEKVSDPEQPLPEPPRGRLOHLQASIS 180
DB 180 PALGRAPWPLGRASSASTPKEPGTGVPSFAEKVSDPEQPLPEPPRGRLOHLQASIS 239
QY 181 QRLGSLDPEMLQCHSEVDFPLGAPKACPDLAGSESSQLIPESAVLPGAGSGQPEAS 240
DB 240 QRLGSLDPEMLQCHSEVDFPLGAPKACPDLAGSESSQLIPESAVLPGAGSGQPEAS 299
QY 241 AFOEVSIRVGSPPSSSGGKKRMNEPWPSPAQVOQESSQAPPEEGAGAVAEEDPPG 300
DB 300 AFOEVSIRVGSPPSSSGGKKRMNEPWPSPAQVOQESSQAPPEEGAGAVAEEDPPG 359
QY 301 EPVOAQPPOPCSSPNRYHGLSPSSQARAAGKAGTAPLHPRLARHDRGNVYRLNMQ 360
DB 360 EPVOAQPPOPCSSPNRYHGLSPSSQARAAGKAGTAPLHPRLARHDRGNVYRLNMQ 419
QY 361 KHYVRRALRSRLRQAMQKRRKKGCGCGGATYTTGSCFLNEQDPHMAAOCPPRA 420
DB 420 KHYVRRALRSRLRQAMQKRRKKGCGCGGATYTTGSCFLNEQDPHMAAOCPPRA 479
QY 421 SEEDTAVGEPPLVPSFQPVPEVSLDPTLPLYSIGSPQGLAETPAEVQALQGLQHA 480
DB 480 SEEDTAVGEPPLVPSFQPVPEVSLDPTLPLYSIGSPQGLAETPAEVQALQGLQHA 539
QY 481 FRPGQERAVMRILSGISTLLVPTGAGKSLCYQLPALLYRRSPCLTLVPSLIMDDQ 540
DB 540 FRPGQERAVMRILSGISTLLVPTGAGKSLCYQLPALLYRRSPCLTLVPSLIMDDQ 599
QY 541 VSGLPPLCTKACIHSQWTRKQRESVLOKRAQVHVLMPLPALVGAAGLPRAAOLPPVA 600
DB 600 VSGLPPLCTKACIHSQWTRKQRESVLOKRAQVHVLMPLPALVGAAGLPRAAOLPPVA 659
QY 601 FACIDEAHCLQWNSHNPFCYLKVRCKVLRBMGVHCFGLTATARTTASDAVQHLAAVE 660

DB 660 FACIDEAHCLQWNSHNPFCYLKVRCKVLRBMGVHCFGLTATARTTASDAVQHLAAVE 719
QY 661 EPDLHGPAVPPTNLHLSVSMRDTPDQALTLTLOGKFPONLSIIYCNREDTERIALL 720
DB 720 EPDLHGPAVPPTNLHLSVSMRDTPDQALTLTLOGKFPONLSIIYCNREDTERIALL 779
QY 721 RTCLHAANVPSSGGAPRTTAAVYHAGCSERRRRVQAFMOGOLR----- 766
DB 760 RTCLHAANVPSSGGAPRTTAAVYHAGCSERRRRVQAFMOGOLRGGGGLWDGAP 839
QY 767 -VVAATVAFGKGLRDVRAVLHLGLPSFSSYQAVRAGRDGPACHLFLQF----- 820
DB 840 ARCAAGCAAGAA---PKLRE-IRAGRP-----GRAPLAPPPTLPYLEPPVNH 886
QY 821 -----QGEPLRELRHVHSDTDFLAVKRLVQVFPACTCTCRPSBOGAVGGERP 873
DB 887 TLMKVALQGEEDLRERRVHSDTDFLAVKRLVQVFPACTCTCRPSBOGAVGGERP 946
QY 874 VPKYPOEAEQLSHQADGPPRCVCGHERALPIQLTVQALDMPER----- 918
DB 947 VPKYPOEAEQLSHQADGPPRCVCGHERALPIQLTVQALDMPEREGEVSHRGVGLS 1006
QY 919 -----AIEYLCTLELHPHMLLATTYTHCRLNCPGPAQLOALAHRCPPILAV 968
DB 1007 PRPLSPALPAIETLLCYLELHPHMLLATTYTHCRLNCPGPAQLOALAHRCPPILAV 1066
QY 969 CLAQOLPEBDPGQSSSVFEDWKLVDNSGWEIASVRLACOLQMHPEPTGVRGTVLV 1028
DB 1067 CLAQOLPEBDPGQSSSVFEDWKLVDNSGWEIASVRLACOLQMHPEPTGVRGTVLV 1126
QY 1029 EPESLAFHLRSPGDLTAAEKDQICDPLVGRVQARERQALRLRRTFOAFHSAVPSCGPC 1088
DB 1127 EPESLAFHLRSPGDLTAAEKDQICDPLVGRVQARERQALRLRRTFOAFHSAVPSCGPC 1186
QY 1089 LEQODEERSTRKDLIGRYFEESEQEPQMGEDAGPEPGQARLODMEDQVCDIRQFLS 1148
DB 1187 LEQODEERSTRKDLIGRYFEESEQEPQMGEDAGPEPGQARLODMEDQVCDIRQFLS 1246
QY 1149 LRPBEKFSRANVARLFGIGSPCYPAQVYGQRRFMWKYLIHSFALVGLATEELLOVAR 1208
DB 1247 LRPBEKFSRANVARLFGIGSPCYPAQVYGQRRFMWKYLIHSFALVGLATEELLOVAR 1306

RESULT 7
ADD01202
ID ADD01202 standard; protein, 1144 AA.
XX
AC ADD01202;
XX
DT 01-JUN-2004 (first entry)
XX
DE Human nucleic acid-associated protein NAAP-40 SEQ ID NO:40.
XX
KW human; nucleic acid-associated protein; NAAP; cytosolic;
KW antiarteriosclerotic; anti-HIV; anti-allergic; cerebroprotective;
KW antiparionomian; anticonvulsant; nootropic; neuroprotective;
KW antiinflammatory; ophthalmological; thyromimetic; antiarthritic;
KW hepatotropic; antibacterial; virucide; protozoacide; antiparasitic;
KW fungicide; gene therapy; cell proliferative disease; cancer;
KW atherosclerosis; hepatitis; neurological disorder; Parkinson's disease;
KW Alzheimer's disease; stroke; epilepsy; developmental disorder;
KW renal tubular acidosis; anaemia; glaucoma; hypochyroidism;
KW autoimmune disorder; inflammatory disorder; AIDS; allergy;
KW atopic dermatitis; arthritis; infection.
XX
OS Homo sapiens.
XX
PN MO2003054219-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-US041115.

XX 19-DEC-2001; 2001US-034306AP.
PR 11-JAN-2002; 2002US-0347633P.
PR 25-JAN-2002; 2002US-0351749P.
PR 22-FEB-2002; 2002US-0359498P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Azimzai Y, Baughn MR, Becha SD, Borowsky ML, Burford N,
PI Elliott VS, Emetling BW, Forsythe LJ, Gotvad AE, Griffin JA,
PI Kadle AE, Knaie R, Lai PG, Lee EA, Lee SY, Li JX, Margulis JP,
PI Rankumar J, Richardson TW, Sprague WW, Swarnakar A, Tang YT,
PI Chawla NK, Warren BA, Yue H;
XX
XX WPI: 2003-559157/52.
DR N-PSDB; ADD01239.
XX
XX New human nucleic acid-associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression, e.g. cancer, hepatitis, AIDS, atherosclerosis,
PT infections.
XX
XX Claim 1; SEQ ID NO 40; 405bp; English.
XX
XX The present invention describes human nucleic acid-associated proteins
CC designated NAAP-1 to NAAP-57. The human NAAPs have cytosolic,
CC antiarteriosclerotic, anti-HIV, antiallergic, cerebroprotective,
CC antiparkinsonian, anticonvulsant, nootropic, neuroprotective,
CC antiinflammatory, ophthalmological, thymometric, antiahrthritic,
CC hepatotropic, antibacterial, virucide, protozoacide, antiparasitic and
CC fungicide activities, and can be used in gene therapy. The NAAP protein
CC and polynucleotide sequences can be used in diagnosing, treating and
CC preventing diseases or conditions associated with the decreased
CC expression or overexpression of NAAP, such as cell proliferative diseases
CC (e.g. cancer, atherosclerosis, hepatitis), neurological disorders
CC (Parkinson's disease, Alzheimer's disease, stroke, epilepsy),
CC developmental disorders (renal tubular acidosis, anaemia, glioma),
CC hypothyroidism), autoimmune/inflammatory disorders (AIDS, allergies,
CC atopic dermatitis, arthritis) and infections (e.g. bacterial, viral,
CC parasitic, protozoal, fungal). The present sequence represents human NAAP
CC -40, from the present invention.
XX
XX Sequence 1144 AA;
SQ
Query Match 94.2%; Score 6050; DB 7; Length 1144;
Best Local Similarity 94.7%; Pred. No. 0;
Matches 1144; Conservative 0; Mismatches 0; Indels 64; Gaps 1;
QY 1 MERLDVVERLQAMERARFRORGRPPSODDYEAAPETTRALYREYRTLKRTTGAQGGGR 60
Db 1 MERLDVVERLQAMERARFRORGRPPSODDYEAAPETTRALYREYRTLKRTTGAQGGGR 60
QY 61 SSELPAABEAPRCWGPMLNRAATKSPQTPGRSRQGSVPDYGRLKANKLTQAG 120
Db 61 SSELPAABEAPRCWGPMLNRAATKSPQTPGRSRQGSVPDYGRLKANKLTQAG 120
QY 121 PALGRRPWPLGRASSTKSTPKPCPGPYPSFAEKVSDPPQLPEQPAPRGLQHLQASLS 180
Db 121 PALGRRPWPLGRASSTKSTPKPCPGPYPSFAEKVSDPPQLPEQPAPRGLQHLQASLS 180
QY 121 PALGRRPWPLGRASSTKSTPKPCPGPYPSFAEKVSDPPQLPEQPAPRGLQHLQASLS 180
Db 121 PALGRRPWPLGRASSTKSTPKPCPGPYPSFAEKVSDPPQLPEQPAPRGLQHLQASLS 180
QY 181 QRLGSLDPGWLQRCHEVVDLFGAPKACRPDLGSEBSQLLPGESAVLGPAGSGSGPEAS 240
Db 181 QRLGSLDPGWLQRCHEVVDLFGAPKACRPDLGSEBSQLLPGESAVLGPAGSGSGPEAS 240
QY 241 AFOEVSIRVSGPQSSSGEKRNNNEEFWESPAQVOQSSQAGPPSEGAVAVEEDPG 300
Db 241 AFOEVSIRVSGPQSSSGEKRNNNEEFWESPAQVOQSSQAGPPSEGAVAVEEDPG 300
QY 301 EPVQAQPPQPCSSSNRPYHGLSPSSQARAKABETAPLHFPRLARDHNGYVRLNMKQ 360
Db 301 EPVQAQPPQPCSSSNRPYHGLSPSSQARAKABETAPLHFPRLARDHNGYVRLNMKQ 360
QY 361 KHVYRGRLARSLRLRKQAMKQKWRKKGCFCGGGATVTTKSCFLNBEFDHMAAQCPRPA 420

Db 361 KHVYRGRLARSLRLRKQAMKQKWRKKGCFCGGGATVTTKSCFLNBEFDHMAAQCPRPA 420
QY 421 SEEDTDAVGEPELVPSQPVPVEVSLDPVTVPLTSLSPSGQLAETPAVEFQALBOLGHOA 480
Db 421 SEEDTDAVGEPELVPSQPVPVEVSLDPVTVPLTSLSPSGQLAETPAVEFQALBOLGHOA 480
QY 481 FRPQGERAVNRIISGISTLLVLPAGKSLCYQLPALLYSRRPCLTVVSPILSMDQ 540
Db 481 FRPQGERAVNRIISGISTLLVLPAGKSLCYQLPALLYSRRPCLTVVSPILSMDQ 540
QY 541 VSGLPCCAKACIHSGMTKORRESVLQKIRAOQVHMLPBEALVGAAGPAPPAQLPVVA 600
Db 541 VSGLPCCAKACIHSGMTKORRESVLQKIRAOQVHMLPBEALVGAAGPAPPAQLPVVA 600
QY 601 FACIDEAHCLSQMSHNRPCYLAVCYLREBMVHCFGLTATATRRRTASDVAOHLVAE 660
Db 601 FACIDEAHCLSQMSHNRPCYLAVCYLREBMVHCFGLTATATRRRTASDVAOHLVAE 660
QY 661 EPDLHGAPVPTNHLISVSMRDTDOALLTLQGRKQNLDSIIYYCNREDTERIALAL 720
Db 661 EPDLHGAPVPTNHLISVSMRDTDOALLTLQGRKQNLDSIIYYCNREDTERIALAL 720
QY 721 RTCLHAAMVPSGGRAPPTTAAVHAGMCSRERRVQRAFMQGLRVVATVAFGMGLDR 780
Db 721 RTCLHAAMVPSGGRAPPTTAAVHAGMCSRERRVQRAFMQGLRVVATVAFGMGLDR 780
QY 781 PDVRAVHLGLPSPBSYVQAVGRAGRDQPAHCHLFLQPOGEDLRELRHVADSTDFL 840
Db 781 PDVRAVHLGLPSPBSYVQAVGRAGRDQPAHCHLFLQPOGEDLRELRHVADSTDFL 840
QY 841 AVKRLVQRPACTCTCTRPSPSEBGAVGERRPVKTPQGEALDSHOAAGPRVCKGH 900
Db 841 AVKRLVQRPACTCTCTRPSPSEBGAVGERRPVKTPQGEALDSHOAAGPRVCKGH 900
QY 901 ERALPIQLTVQALMPBEALJETTLCYELRPHHMLLELTATYTHCRINCPGPAQOLALA 960
Db 901 ERALPIQLTVQALMPBEALJETTLCYELRPHHMLLELTATYTHCRINCPGPAQOLALA 960
QY 961 HRCBPPLAVCLAQQLPBDPGQSSSVFEDMYLVDSMGMELASVRRALCOLQMDHEPRTGV 1020
Db 961 HRCBPPLAVCLAQQLPBDPGQSSSVFEDMYLVDSMGMELASVRRALCOLQMDHEPRTGV 1020
QY 1021 RRGTVLVERSBLAFHLSRQDLTAEEKDQICDPLYGRVQAREQALRLRRTQAHSV 1080
Db 1021 RRGTVLVERSBLAFHLSRQDLTAEEKDQICDPLYGRVQAREQALRLRRTQAHSV 1080
QY 1081 AFPSGCPCLEQDERSTRKDLIGRYFEEBEGQEPGMDAOGPEBGOARLODMEDQVR 1140
Db 1081 AFPSGCPCLEQDERSTRKDLIGRYFEEBEGQEPGMDAOGPEBGOARLODMEDQVR 1140
QY 1141 CDIROPFLSRPEEKSSRAVARIFHGIGSPCYAQQVYGODRRFRKTYLHLSFHALVGLAT 1200
Db 1141 CDIROPFLSRPEEKSSRAVARIFHGIGSPCYAQQVYGODRRFRKTYLHLSFHALVGLAT 1200
QY 1201 BELLOVAR 1208
Db 1201 BELLOVAR 1208
QY 1137 BELLOVAR 1144
Db 1137 BELLOVAR 1144
RESULT 8
ABM84821
ID ABM84821 standard; protein; 1142 AA.
XX ABM84821;
AC
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:5070.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dthp.
XX
OS Homo sapiens.

XX MO2004023973-A2.
 XX 25-MAR-2004.
 XX 12-SEP-2003; 2003WO-US028227.
 XX 12-SEP-2002; 2002US-04102595.
 XX 12-SEP-2002; 2002US-04102605.
 XX (INCY-) INCYTE CORP.
 XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,
 XX Harshbome TA, Suchorski MT, Altus CM, Plets SJ, Rider LV,
 XX Mooney EM, Deleage AM, Panesar IS, Banville SC, Reddy TP,
 XX Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerlein EH,
 XX Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL,
 XX Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vilt UA, Klinton ES,
 XX Xu Y, Kwong M, Policky CJ, Hurwitz BL, Ma Y, Jackson JL, Gietzen D,
 XX Patuary S, Shi X, Suarez CJ,
 XX WPI; 2004-329368/30.
 XX N-PSDB; ACN43473.
 XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 XX in diagnosing a condition, disease or disorder associated with human
 XX molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 XX in gene mapping.
 XX Claim 27, Page; 190pp; English.
 XX The invention relates to novel diagnostic and therapeutic polynucleotides
 XX selected from one of the 2722 sequences defined in the specification. A
 XX polynucleotide of the invention may have a use in gene therapy. The human
 XX diagnostic and therapeutic polynucleotides (dntp) or polypeptides may be
 XX used to diagnose a particular condition, disease or disorder associated
 XX with human molecules, e.g. cell proliferative disorders,
 XX autoimmune/inflammatory disorder, developmental disorder, endocrine
 XX disorder, neurological disorder, gastrointestinal disorders, or
 XX infections caused by virus, bacteria, fungi or parasite. The dntp
 XX molecules may also be used in genetic mapping, in identifying individuals
 XX from minute biological samples, in detecting single nucleotide
 XX polymorphisms, as molecular weight markers, and for somatic or germline
 XX gene therapy. The present sequence represents a dntp protein of the
 XX invention. Note: The sequence data for this patent is not represented in
 XX the printed specification, but was obtained in electronic format directly
 XX from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX Sequence 1142 AA;
 XX Query Match 93.9%; Score 6029; DB 8; Length 1142;
 XX Best Local Similarity 94.5%; Pred. No. 0;
 XX Matches 1141; Conservative 0; Mismatches 1; Indels 66; Gaps 1;

QY 301 EPVQAPQPCSSPSNPRYHGLSPSSQARAKAGTAPLHLPRLARRDNGYVLANKKO 360
 DB 301 EPVQAPQPCSSPSNPRYHGLSPSSQARAKAGTAPLHLPRLARRDNGYVLANKKO 360
 QY 361 KHYVGRALRSRLRKQAMKQKMKKGCFCGGGATTTTKSSCFINQEPDHAACPRPA 420
 DB 361 KHYVGRALRSRLRKQAMKQKMKKGCFCGGGATTTTKSSCFINQEPDHAACPRPA 420
 QY 421 SEEDTDAVGPPIVSPQVPEVPSLDPTVLPLVSLGSGGQLAETPAVFOALBQGHQA 480
 DB 421 SEEDTDAVGPPIVSPQVPEVPSLDPTVLPLVSLGSGGQLAETPAVFOALBQGHQA 480
 QY 481 FRPQGERAVMBILSGISTLVLPAGKSLCYQLPALYSRRSPCLTVVSPLSLMDQ 540
 DB 481 FRPQGERAVMBILSGISTLVLPAGKSLCYQLPALYSRRSPCLTVVSPLSLMDQ 540
 QY 541 VSGLPPLCKAACIHSQMTKRORESYLOKIRAAQVHVMLTPEALVAGGLPPAOLPPVA 600
 DB 541 VSGLPPLCKAACIHSQMTKRORESYLOKIRAAQVHVMLTPEALVAGGLPPAOLPPVA 600
 QY 601 FACIDBAHCLSGWSNFRPCYLKVCVLRERGVHCFGLTATVTRTASDVQCHLVAE 660
 DB 601 FACIDBAHCLSGWSNFRPCYLKVCVLRERGVHCFGLTATVTRTASDVQCHLVAE 660
 QY 661 EPDLGPAVPPTNLHLSVMDTDQALLTLQGRFQNDIITTCNRREDTERIAALL 720
 DB 661 EPDLGPAVPPTNLHLSVMDTDQALLTLQGRFQNDIITTCNRREDTERIAALL 720
 QY 721 RTCLHAAMVPGSGGAPKTTAAVYHAGCRRRRRVRORAFOGQLRVVATVAFGNGLDR 780
 DB 721 RTCLHAAMVPGSGGAPKTTAAVYHAGCRRRRR----- 755
 QY 781 PIVRAVLHGLPSPRESYVQVGRAGRDQPAHCHLFLQPOGEDLRELRRHYHADTDL 840
 DB 781 PIVRAVLHGLPSPRESYVQVGRAGRDQPAHCHLFLQPOGEDLRELRRHYHADTDL 840
 QY 756 -----GEDIRELRRHYHADTDL 774
 DB 756 -----GEDIRELRRHYHADTDL 774
 QY 841 AVKRLVQVRFPACTCTCTPRPSEOGAVGGERPVKYPQABEOLSHQAAPRRVCMGH 900
 DB 841 AVKRLVQVRFPACTCTCTPRPSEOGAVGGERPVKYPQABEOLSHQAAPRRVCMGH 900
 QY 775 AKRLVQVRFPACTCTCTPRPSEOGAVGGERPVKYPQABEOLSHQAAPRRVCMGH 834
 DB 775 AKRLVQVRFPACTCTCTPRPSEOGAVGGERPVKYPQABEOLSHQAAPRRVCMGH 834
 QY 901 EBALPQLTVQALDMPBEAIEITLCLYLHPHMLLATTYTHCLNCPGAPAOIALA 960
 DB 901 EBALPQLTVQALDMPBEAIEITLCLYLHPHMLLATTYTHCLNCPGAPAOIALA 960
 QY 835 EBALPQLTVQALDMPBEAIEITLCLYLHPHMLLATTYTHCLNCPGAPAOIALA 894
 DB 835 EBALPQLTVQALDMPBEAIEITLCLYLHPHMLLATTYTHCLNCPGAPAOIALA 894
 QY 961 HRCPLAVLTAOLPEDPCQSSSVFDMVKLVDSMGWLAIVRRALCOLQMDHERTEY 1020
 DB 961 HRCPLAVLTAOLPEDPCQSSSVFDMVKLVDSMGWLAIVRRALCOLQMDHERTEY 1020
 QY 895 HRCPLAVLTAOLPEDPCQSSSVFDMVKLVDSMGWLAIVRRALCOLQMDHERTEY 954
 DB 895 HRCPLAVLTAOLPEDPCQSSSVFDMVKLVDSMGWLAIVRRALCOLQMDHERTEY 954
 QY 1021 RRGTVLVERSELAFILRSRGLTAEKQICDPLVGRVQAREBROLARRTPQAFHSV 1080
 DB 1021 RRGTVLVERSELAFILRSRGLTAEKQICDPLVGRVQAREBROLARRTPQAFHSV 1080
 QY 955 RRGTVLVERSELAFILRSRGLTAEKQICDPLVGRVQAREBROLARRTPQAFHSV 1014
 DB 955 RRGTVLVERSELAFILRSRGLTAEKQICDPLVGRVQAREBROLARRTPQAFHSV 1014
 QY 1081 AFPSGCPCLQODDEKSTRKLKLAGRYFEESGOBEGMEDAQGPFGAQLQDMDQYR 1140
 DB 1081 AFPSGCPCLQODDEKSTRKLKLAGRYFEESGOBEGMEDAQGPFGAQLQDMDQYR 1140
 QY 1015 AFPSGCPCLQODDEKSTRKLKLAGRYFEESGOBEGMEDAQGPFGAQLQDMDQYR 1074
 DB 1015 AFPSGCPCLQODDEKSTRKLKLAGRYFEESGOBEGMEDAQGPFGAQLQDMDQYR 1074
 QY 1141 CDIRQTLSTRPEKSSRAVVAIFGIGSPCYPAQYGGDRFRMKYLHSPALVGLAT 1200
 DB 1141 CDIRQTLSTRPEKSSRAVVAIFGIGSPCYPAQYGGDRFRMKYLHSPALVGLAT 1200
 QY 1075 CDIRQTLSTRPEKSSRAVVAIFGIGSPCYPAQYGGDRFRMKYLHSPALVGLAT 1134
 DB 1075 CDIRQTLSTRPEKSSRAVVAIFGIGSPCYPAQYGGDRFRMKYLHSPALVGLAT 1134
 QY 1201 EELLQAVAR 1208
 DB 1201 EELLQAVAR 1142
 DB 1135 EELLQAVAR 1142

RESULT 9
 AAB2095
 ID AAB2095 standard; protein; 756 AA.
 AC AAB2095;
 XX
 DT 11-DEC-2000 (first entry)

XX DE Human RecQ4 helicase mutant, mut-2.
 XX KW RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;
 KW poikiloderma congenitale; autosomal recessive; skin disorder;
 KW dermatology; antibody; prenatal diagnosis; gene therapy; mutant; mutain.
 OS Homo sapiens.
 XX MO200043522-A1.
 XX PD 27-JUL-2000.
 XX PF 19-JAN-2000; 2000WO-JP000233.
 XX PR 19-JAN-1999; 99JP-00011218.
 XX PA (AGEN-) AGENE RES INST CO LTD.
 XX PI Kitao S, Shimamoto A, Furuchi Y;
 XX WP1; 2000-524241/47.
 XX DR N-PSDB; AAA72365.
 XX PT RecQ4 helicase gene, gene products and antibody, used in the diagnosis
 XX and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.
 XX PS Example 3; Page; 115pp; Japanese.
 XX CC The present sequence represents a mutant human RecQ4 helicase, mut-2. The
 CC cDNA encoding sequence contains a C to T substitution relative to the
 CC cDNA encoding the wild-type RecQ4 helicase (AAA72321), which changes a
 CC Gln codon (CAG) to a stop codon (TGA), causing premature truncation of
 CC the encoded protein. The invention relates to the genomic DNA sequence of
 CC human RecQ4 helicase (AAA72320). Mutations in this gene, located on
 CC chromosome 8q24.3, are the cause of Rothmund-Thomson syndrome (also known
 CC as poikiloderma congenitale), an autosomal recessive skin disorder
 CC principally occurring in females and often accompanied by juvenile
 CC cataracts, saddle nose, congenital bone defects, hypogonadism and
 CC disturbances in the growth of hair, nails and teeth. The invention also
 CC relates to vectors and host cells comprising the human RecQ4 helicase
 CC genomic sequence. It additionally encompasses use of the RecQ4 helicase
 CC protein as a therapeutic and anti-RecQ4 antibodies as diagnostic agents.
 CC The RecQ4 helicase gene and its products, and anti-RecQ4 helicase
 CC antibodies are useful in the diagnosis, especially prenatal diagnosis,
 CC and treatment of Rothmund-Thomson syndrome. The genomic sequence may
 CC especially be used in gene therapy for this condition. Note: The present
 CC sequence is not shown in the specification, but is derived from the wild-
 CC type human RecQ4 helicase shown on pages 83-92
 XX SQ Sequence 756 AA:
 Query Match 62.4%; Score 4006; DB 3; Length 756;
 Best Local Similarity 100.0%; Pred. No. 2.6e-284;
 Matches 756; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MERLDVARERIQAMERARRRQGRPPSODDYEAAPETRALYREYRTTKRTTGAAGGIR 60
 DB 1 MERLDVARERIQAMERARRRQGRPPSODDYEAAPETRALYREYRTTKRTTGAAGGIR 60
 QY 61 SSESIPAAAEAPBERCWPGLHNRATKSPQPTPERSQGVVDYQGRILKNTLKTQLQAG 120
 DB 61 SSESIPAAAEAPBERCWPGLHNRATKSPQPTPERSQGVVDYQGRILKNTLKTQLQAG 120
 QY 121 PALGRRPWPLGRASSKASTPKPGTGVPSFAEKVSDPEPOLPEPQPRGRLOHLQASIS 180
 DB 121 PALGRRPWPLGRASSKASTPKPGTGVPSFAEKVSDPEPOLPEPQPRGRLOHLQASIS 180
 QY 181 QRLSLDPGWTQRCHESEVDTLGAAPKACRPDLGSEESOLLIPGESAVLGPAGSGSPBAS 240
 DB 181 QRLSLDPGWTQRCHESEVDTLGAAPKACRPDLGSEESOLLIPGESAVLGPAGSGSPBAS 240
 QY 241 AFGESVIRVSGPSPSSSGGKRWMEEPWESPAPQVQSSQAGPSSBAGAVVAEEDPPG 300

DB 241 AFGESVIRVSGPSPSSSGGKRWMEEPWESPAPQVQSSQAGPSSBAGAVVAEEDPPG 300
 QY 301 EPVQAQPPQPCSSPNPRYHGLSPSSQARAGAGTAPLHFPPLARHGRGNVYRLNMKQ 360
 DB 301 EPVQAQPPQPCSSPNPRYHGLSPSSQARAGAGTAPLHFPPLARHGRGNVYRLNMKQ 360
 QY 361 KHYVGRALRRLRKQAMKQMKRKGCCFGGGAIVTTKSCFLNBFQDMAAQCPRPA 420
 DB 361 KHYVGRALRRLRKQAMKQMKRKGCCFGGGAIVTTKSCFLNBFQDMAAQCPRPA 420
 QY 421 SEEDTDVAGPEPLVPSPQVEVPSLDPVTYLPLYSIGPSGQLAETPAVFOALSGHQ 480
 DB 421 SEEDTDVAGPEPLVPSPQVEVPSLDPVTYLPLYSIGPSGQLAETPAVFOALSGHQ 480
 QY 481 FRPQGERAVMRILSGISTLLVLTPTGAGKSLCYQDPALLYSRRSPCLTLVSPLLSMDQ 540
 DB 481 FRPQGERAVMRILSGISTLLVLTPTGAGKSLCYQDPALLYSRRSPCLTLVSPLLSMDQ 540
 QY 541 VSGLPPLCKAACHSGMTRKQRESVLQKIRAAQVHVMLTREBALVGAGLPPAAQLPPVA 600
 DB 541 VSGLPPLCKAACHSGMTRKQRESVLQKIRAAQVHVMLTREBALVGAGLPPAAQLPPVA 600
 QY 601 PACIDEAHCLSQMSHNPFCYLRYCKVLRBMGYHCFGLTATATRTTASDVAQHLAVAE 660
 DB 601 PACIDEAHCLSQMSHNPFCYLRYCKVLRBMGYHCFGLTATATRTTASDVAQHLAVAE 660
 QY 661 EPDLHGPAVPTNLHLVSMDRDTDOALLTLLOGKRFPQNLDSIIYYCNRREDTERIAL 720
 DB 661 EPDLHGPAVPTNLHLVSMDRDTDOALLTLLOGKRFPQNLDSIIYYCNRREDTERIAL 720
 QY 721 RTCLHAAMVPSGGRAPTTKEATHAGCSEHRRRV 756
 DB 721 RTCLHAAMVPSGGRAPTTKEATHAGCSEHRRRV 756
 RESULT 10
 ID ADW44021 standard; protein; 1216 AA.
 AC ADW44021;
 DT 24-MAR-2005 (first entry)
 DE Mouse RECQL4 protein.
 KW RECQL4; helicase; protein deactivation; growth disorder;
 KW Rothmund-Thomson syndrome; skin disorder; bone disease;
 KW musculoskeletal disease; enzyme.
 OS Mus musculus.
 PN WO2005001085-A1.
 PD 06-JAN-2005.
 XX 25-JUN-2004; 2004WO-JP009380.
 XX 27-JUN-2003; 2003JP-00185409.
 XX (AGEN) NAT INST RADIOLOGICAL SCT.
 XX Abe M;
 DR WP1; 2005-081681/09.
 XX N-PSDB; ADW44020.
 XX Novel non-human mammal, preferably rodent such as mouse exhibiting
 PT characteristics of Rothmund-Thomson syndrome, being RECQL4-gene lacking
 PT mouse having mutation in RECQL4-gene, useful as model of human Rothmund-
 PT Thomson syndrome.
 XX Disclosure; SEQ ID NO 2; 69pp; Japanese.

XX This invention describes a novel knock-out mouse which lacks the RECQ4 gene or has a mutation in one of exons 13 to 22 of the RECQ4 gene resulting in a loss of helicase activity. The mouse described in the invention exhibits the characteristics of Rothmund-Thomson syndrome and is suitable as a model of human Rothmund-Thomson syndrome. The mouse exhibits growth delay, skin abnormality and bone-formation defects similar to human Rothmund-Thomson syndrome. This sequence represents a murine RECQ4 which is missing in mice exhibiting Rothmund-Thomson syndrome.

XX Sequence 1216 AA;

61.4%; Score 3944; Db 9; Length 1216;

Query Match Best Local Similarity 63.9%; Pred. No. 1,7e-279; Matches 795; Conservative 110; Mismatches 267; Indels 72; Gaps 13;

```

QY 1 MERLADVERLQAWERARFRQRRPSQDDVEAAEEETRALYREYRLKRTTGAAGGILR 60
DB 1 MERLATVABRLQEWERAFAHLHGRPAKGVAAAEETRALYREYRLKQAVRQADNHR 60
QY 61 SSE-SLPAAAEAPPCPCGHNAATKSPOPTGRSQSVDPYQGLKANKTIOA 119
DB 61 VLEQSLAAAEBAEAPSCWGPPLSRAATQNTQSMKQSLSSVDQGRKLANKNTTQT 120
QY 120 GPALGRRRWPLGRASSKASTPKPGTGPVSPAEKVSPEPPOLPBPORPGRLOHQA 179
DB 121 GPTQSRKQLOKRSLSVTPAPRPSPKTESPCPDADDLAPRPFRPLQLOOLRSSL 180
QY 180 SQRLSLDPGWLQRCHEVDPFLGAPKACRPDLGSEBSQLIPGSAVLGPGAGS----Q 235
DB 181 SRRLLSLDPGWLERCHNRVSDLEVPAGCGLDLSAEESQPMQSGKVNLAIDPIQSEVSQ 240
QY 236 GPESAFQEVSTIRVSGPOSSSGEKRWNEEPWESPAVOQESSQAPPSGAGAVAE 295
DB 241 SPEAIAQCPAQLSOSPXSINSKGRKKNK-GBDFQ-DQPSGAGLSPGAATVHG 298
QY 296 BDPPEBPQAOCPPOCSSPNRHYGLSPSQARAGKAGTAPLHPPLARHDRGVYR 355
DB 299 QDPPEBPQVNVPPQNS-----SNQARTEKAGITHLASPPASLDKGNVIR 347
QY 356 LMMKQHYVGRALSRLLRKQAMQKWKKECEGCGGATVTTESCLNQPDMMAQ 415
DB 348 LMMKQKRFVRVAGANGRLRKQVMQKWKQAAAGSGSPRATDQTCRCQGFHMASQ 407
QY 416 CPRPA-----SEEDTAVGP-----EPLVSPPPVPEVPSL 446
DB 408 CSQPGPTLVQEEGRDDKQPISTLEEVAQRTGASCHHSGETQPAAEELQ-VPHCP- 465
QY 447 DPTVPLVLSGPGSGLAETPAEVPQALBOLGHOAPRPGQERAVMRLSGISTLVLP 506
DB 466 --PMSPELVPPGPGVQAEPAEVPQALBOLGTRAFPGQERAVMRLSGISTLVLP 523
QY 507 GKSICVQALPALTYSRSPCLTVVSPSLIMDQVSGLPCLKAACIHSGMTKORSEVL 566
DB 524 GKSICVQALPALTYSRSPCLTVVSPSLIMDQVSDPCLKAACIHSGMTKORSEVL 583
QY 567 QKIRPAQVHVMLTPEALVGA-----GLPRAAQLPVPVAFACIDEAHCLISQMSH 621
DB 584 KKVRAAQVHVMLTPEALVGA-----GLPRAAQLPVPVAFACIDEAHCLISQMSH 643
QY 622 LWCVYLBERMGVHGLTATARTASDVQOHLVAEERPLHGAPVPTLHLISVSD 681
DB 644 LWCVYLBERMGVHGLTATARTASDVQOHLVAEERPLHGAPVPTLHLISVSD 703
QY 682 RPTDQALTLTQKRFQNLDSIIYCNREDETERIALRTCLHAAMVPSGGRAPKTTA 741
DB 704 RPSDQALTLTQKDRRTLDVITCTRERIQNGMIALLRICLSMWGDSRPGCGEALIA 763
QY 742 EAYHAGMCRERRRRVQAFMGQGLRVVTVAFAGMLDPRVRAVTHLGLPSFESYVA 801
DB 764 EAYHAGMSSQERRRVQAFMRGHLRMVAVTVAFAGMLDPRVRAVTHLGLPSFESYVA 823

```

```

QY 802 VGRAGDQGANHCHLFLQPGSEDLRELRHVHADSTDFLAVKRLVQVFPACTCTTRPP 861
DB 824 IGRAGRGKPAHCHLFWHPQGEDLMELRRHHAHSDTDFLAVKRLVQVFPACTCTTRPP 878
QY 862 SEQSGAVGGERPVKYPQPOAEQSL-HQAPGPRRVCMGERALPIQLTQALMPEAI 920
DB 879 -----QRPVKSSEPEEKESGQTYPVLGQACLGHERALPVOSTVQALMTBEAI 929
QY 921 ETLCTYLEHPHMLLELTATYTHCRINCPGPAQLOALAHRCPLAVCLAAQOLPEDPGQ 980
DB 930 ETLCTYLEHPHMLLELTATYTHCRINCPGPAQLOALAHRCPLAVCLAAQOLPEDPGQ 989
QY 981 GSSSEEDMVNLVDSMGWELASVRAALCOLQWDEHPRTGVRRGTGLVSESLAFHLRSP 1040
DB 990 GRSSLERGVNLEADSMGKTLASVQALHQLKWDPEPKGAAGTGLVVKFSELAFLHLSR 1049
QY 1041 GDLTAERKQICDPLVGRVQAREQALALRTQAFHSVAFPSGCGCLEQODEERSTR 1100
DB 1050 GDLTDEERKQICDPLVGRVQAREQALALHQLKWDPEPKGAAGTGLVVKFSELAFLHLSR 1109
QY 1101 KDLGRYFEESEEGQEPGAMEDAOPPEPQARLOPMEQVRCDIQFLSLRPEKFSRAV 1160
DB 1110 KTLVSYTFEESE-EEETMTDTQGPFGQTOLOPMEQVRCDIQFLSLRPEKFSRAV 1168
QY 1161 ARIFHGIGSPCYPAQVYGODRRFWRKYLHLSFHALVGLATEELL 1204
DB 1169 ARIFHGIGSPCYPAQVYGODRRFWRKYLHLSFHALVGLATEELL 1212

```

RESULT 11
AAB20994
ID AAB20994 standard; protein; 554 AA.
XX
AC AAB20994;
XX
DT 11-DEC-2000 (first entry)
XX
DE Human RecQ4 helicase mutant, mut-1.
XX
KW RecQ4 helicase; human; Rothmund-Thomson syndrome; chromosome 8q24.3;
KW pollioderma congenitale; autosomal recessive; skin disorder;
KW dermatology; antibody; prenatal diagnosis; gene therapy; mutant; mutcin.
OS Homo sapiens.
PN WO200043522-A1.
XX
PD 27-JUL-2000.
XX
PF 19-JAN-2000; 2000WO-JP000233.
XX
PR 19-JAN-1999; 99JP-00011218.
XX
PA (AGEN-) AGENE RES INST CO LTD.
XX
PI Kitao S, Shimamoto A, Furuichi Y;
XX
DR WPI; 2000-524241/47.
XX
DR N-PSDB; AAA72364.
XX
PT RecQ4 helicase gene, gene products and antibody, used in the diagnosis
PT and treatment of Rothmund-Thomson syndrome, e.g. by gene therapy.
XX
PS Example 3; Page; 115pp; Japanese.
XX
XX The present sequence represents a mutant human RecQ4 helicase, mut-1.
CC This cDNA encoding sequence contains a 7 bp deletion relative to the cDNA
CC encoding the wild-type RecQ4 helicase (AAA72321), which alters the
CC reading frame and causes premature truncation of the encodes protein. The
CC invention relates to the genomic DNA sequence of human RecQ4 helicase
CC (AAA72320). Mutations in this gene, located on chromosome 8q24.3, are the
CC cause of Rothmund-Thomson syndrome (also known as pollioderma
CC congenitale), an autosomal recessive skin disorder principally occurring

in females and often accompanied by juvenile cataracts, saddle nose, congenital bone defects, hypogonadism and disturbances in the growth of hair, nails and teeth. The invention also relates to vectors and host cells comprising the human RecQ4 helicase genomic sequence. It additionally encompasses use of the RecQ4 helicase protein as a therapeutic and anti-RecQ4 antibodies as diagnostic agents. The RecQ4 helicase gene and its products, and anti-RecQ4 helicase antibodies are useful in the diagnosis, especially prenatal diagnosis, and treatment of Rothmund-Thomson syndrome. The genomic sequence may especially be used in gene therapy for this condition. Note: The present sequence is not shown in the specification, but is derived from the wild-type human RecQ4 helicase shown on pages 83-92

Sequence 554 AA;

Query Match 45.5%; Score 2920; DB 3; Length 554;
Best Local Similarity 100.0%; Pred. No. 6.5e-205;
Matches 550; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MERLRDVERIQAWEARARFRQRGRPPSODDVEAARETRALYREYRLKRTTGGAGGLR 60
DB 1 MERLRDVERIQAWEARARFRQRGRPPSODDVEAARETRALYREYRLKRTTGGAGGLR 60
QY 61 SSESIPAAAEAPERPCWGHILNRAATKSPQTPGRSRQSVDPYQRLKANTLKGTLQAG 120
DB 61 SSESIPAAAEAPERPCWGHILNRAATKSPQTPGRSRQSVDPYQRLKANTLKGTLQAG 120
QY 121 PALGRPPPLGRASKASTPYRPGTGPVPSFAEKVSDPEPQPEPQPRGRLQHLQASLS 180
DB 121 PALGRPPPLGRASKASTPYRPGTGPVPSFAEKVSDPEPQPEPQPRGRLQHLQASLS 180
QY 181 QRLSLDGMVLRCHSEVDPDLGAPKACRPDLGSEESQLLTPEGSAVLGPRAGSGGPEAS 240
DB 181 QRLSLDGMVLRCHSEVDPDLGAPKACRPDLGSEESQLLTPEGSAVLGPRAGSGGPEAS 240
QY 241 AFOEVSIRVSGPOSSSGEGRKRWNEEPWESPAVOQOQSSQAGPSEGAANAVEDPBG 300
DB 241 AFOEVSIRVSGPOSSSGEGRKRWNEEPWESPAVOQOQSSQAGPSEGAANAVEDPBG 300
QY 301 EPVQAOPPOPCSSPSPNPRYHGLSPSSQARAGKAGTADLPIPRILARHNRGNVYRLNKKQ 360
DB 301 EPVQAOPPOPCSSPSPNPRYHGLSPSSQARAGKAGTADLPIPRILARHNRGNVYRLNKKQ 360
QY 361 KHYVRGRLRSLRLRKQWKQKMKKRCRGCGATVTTTKESCPLNEPDMAAQCPPEPA 420
DB 361 KHYVRGRLRSLRLRKQWKQKMKKRCRGCGATVTTTKESCPLNEPDMAAQCPPEPA 420
QY 421 SSEDTDVAGPEPLVSPQVPEVPSLDPTVLPLYSIGSGQLAETPAVFOALBQLGHQA 480
DB 421 SSEDTDVAGPEPLVSPQVPEVPSLDPTVLPLYSIGSGQLAETPAVFOALBQLGHQA 480
QY 481 FRPQOEBAVMRIISGISTLLVLTGAGKSLCYQPLALYSRRSPCLTLVSPLLSLMDMQ 540
DB 481 FRPQOEBAVMRIISGISTLLVLTGAGKSLCYQPLALYSRRSPCLTLVSPLLSLMDMQ 540
QY 541 VSGLPCLCKA 550
DB 541 VSGLPCLCKA 550

```

RESULT 12

ABB63317 ID ABB63317 standard; protein; 1579 AA.

AC ABB63317;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 16743.

KM Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKB) PE CORP NY.

XX Venter JC, Adams M, Li FWD, Myers EW,

XX WPI; 2001-656860/75.

XX N-PSDB; ABL07420.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

XX Disclosure; SEQ ID NO 16743; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-ABBS72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 1579 AA;

Query Match 26.8%; Score 1723.5; DB 4; Length 1579;
Best Local Similarity 28.9%; Pred. No. 1.3e-116;
Matches 467; Conservative 217; Mismatches 459; Indels 471; Gaps 45;

```

QY 10 RLQAWERAPFRQRGRPPSODDVEAARETRALYREYRLKRT----- 51
DB 16 RLVKWERDQFKKQKGRVPSKYDIRDSQERDSYMYKLTSLFLETLNDVLSGDYDL 75
QY 52 -TGQAGC-----GLASS--ESLP----- 66
DB 76 EMSQASDPGVSMLDQVSLNBPQLPRDLSALVGPQSSGNLEIFQVSGSFSNLIDLPPN 135
QY 67 -----AAAEAP-EPRCWGHILNRAATKSPQ-----TPGRSRQG 100
DB 136 RQVLTNLNVDENHVIKRFVAVELPIQNQANGLNWKPPAPQVPEASKAGHGKQ- 194
QY 101 SVRPYQGLRLKANTLKTU-QAGPALGR--FWPLGRASKASTPYRPGTGPVPSFAEKVSD 157
DB 195 --PRAQASLKPSSLAKLFOSSRGPAKKNPKKPRKYSRCVSSSSSTTSLSSVPTDHHLELDF 252
QY 158 EPPQLEPPQPRPGRLLQ-----HLQASLSQRLGSLDPLGMLQR-----CHSEVP- 199
DB 253 ETLIIRKAQRYKKEQQAIAANNPMLASHESKESIKT---LYDDGRLRNTYENTLDEVPF 309
QY 200 ----DFVGAPEK-----ACRPDLGSEESQLLIPEBSAVL-----GP 230
DB 310 AEANNNGTSKTKTNFGLANLDSLKLKPTVKEKVLQAKPPQMAIIQELQTDNMSNMNQKP 369
QY 231 GAGSQGEASAFQEVSTRVSGPOSS-----SGEGRKRWNE----- 267
DB 370 DHLNHTPPASSQKSVAPK-NKPPSEGETSDSDSVVAESEBQEPQETRLQSKRRKIVS 428
QY 268 -----PWESPAVOQOES--SQAGP----- 284
DB 429 TASQKVEVAAPVELPKNVEPTEETFAQENDPFSADEQDATTYPENKKQKAKRQAKG 488

```

```

QY 285 -----PSEGAVALVEEDPPEVQAPOP----- 310
DB 489 QUTTKPKAPKTEKKAVALKAKKPKAKKPRNSKKAIVAPAPDEEDKQPLNPEL 548
QY 311 -----CSPSPNPRY-----GLSPSQAPAKAGETAP 338
DB 549 KYVALAEADITSVPRINVOLEADATQARYIRTPAAGPNPGLSEGSNIRYDEKKAAR 608
QY 339 LHIFFPLAHD-RGNVYRLMOKHYRG-RALRSLLKQMKQKR-----KKG 387
DB 609 KKLERRIAAGKLNENVTJNIOCKKVRKGSVNSKYKQQRHKKRVAALSGPDMWG 668
QY 388 ECFGGGATVTTKESCFLEQFDMMAOC-----PRP----- 419
DB 669 GCDGG-----VLT-----CFGGGVHFPQOCKYKGDLSLLPLSAOLEEPPSPPTLAQEMA 721
QY 420 -----ASEEDTAVGPPELVSPQP----- 439
DB 722 SGQAVVAHSRNIIRLPQAAVAAILQGBELNESEEDQESGDEEVQOHDPNMSDEMDV 781
QY 440 -----VPE-----VPELDP-----VL 451
DB 782 DFEALDAVEASISQVSOEKAAPIKTYVGHKIPEEFKQAGLDITPASSNSQHGK 841
QY 452 PLYSLGPSQGLAETPAVEFQALBOLGHOAFRPGQERAVNRILSGISTLVLPAGAKSLC 511
DB 842 PLYDLPPDSGVDTTEBEVLEALMFQHTNPKQODAIKRTLSGLSVLTLGSGSKSLC 901
QY 512 YQPLALLYSRRSPCLTLVSPPLSLMDQVSGLPCLKAACHSGMTKRESVLQKIRA 571
DB 902 YQPLALLYSRRSPCLTLVSPPLSLMDQVSGLPCLKAACHSGMTKRESVLQKIRA 571
QY 572 AQVHTMLTPBALV-----GAGLPPAOLPVVAFACIDBAACLSQMSNPPCTLRVC 625
DB 962 GERDILVSPBEAVAGBRATFGAI--LRQLPRIAFICIDEACVQSWNSFPRSYLMTC 1019
QY 626 KVLRRNGVCFGLPATATRTASDVAOHLAAV--EPDLGAPAPETNHLTVSMDRDT 684
DB 1020 KVLRRNGVCFGLPATATRTATLPRVSIINHLGSDERGILSILPLPDLNLVLSVDENR 1079
QY 685 DQALLTLQGRFQNDLSIIYCNRRREDTERIALNTCLHAAMPVSGGRAPKT-----T 740
DB 1080 DQALLTLQGRFQNDLSIIYCNRRREDTERIALNTCLHAAMPVSGGRAPKT-----T 740
QY 741 AEAHAGMCRERRRQVAPAMOGOLRVYVATVAFGMLDPRVRAVLHLGLPSPFESYVQ 800
DB 1140 AEAHAGMCRERRRQVAPAMOGOLRVYVATVAFGMLDPRVRAVLHLGLPSPFESYVQ 800
QY 801 AVBRAGDQAPACHLPLQOGEDLRBRHVANDTDPLAVKRLVQRPACTCTCTRP 860
DB 1200 EIGRAGDQAPACHLPLQOGEDLRBRHVANDTDPLAVKRLVQRPACTCTCTRP 860
QY 861 PSEOGAVGGERPVKYPQEBOLSHQAPGR-RVCMGHERALPQLTVQALMDPEEA 919
DB 1255 --DKEASKRTALRP-----LEBDGRVHMCQPHREIGSFSEKTEVMMDIPAEN 1300
QY 920 IETLLCYLELHPHMELELATTTTHCLNCPGPAQOIALANRCPPLAVCLAQQLPEDP- 978
DB 1301 IETLLCYLELHPHMELELATTTTHCLNCPGPAQOIALANRCPPLAVCLAQQLPEDP- 978
QY 979 GQGSSTVEPFWKLVDSMGWELASVVRALQOLMDHEPRGVRGTVLVEFSELAHRL 1038
DB 1361 KEBNSNIEFVTDIAAGIGMSGVKQLQDLER--VKVNGYPRKRSITVSPDLGRIRK 1418
QY 1039 SPQDLTAERKQOICDPLYGVARERQALARLRTTFAFHSVAPSGCPLEQO-DEERS 1097
DB 1419 VPGDFTSESIDNALDLYTRSVMKQERTQLOLYAAGLAAYASSGQCCNADFPQDRG 1478
QY 1098 TRKMDLGRYEEBEQO-----BRGMEDAQBEPQOARLQDMEDVQCDIRFSLRP 1151
DB 1479 EQLFAIVANFPANDYPODLELEIENVPD-----ENIIDVHALINMYP 1523
QY 1152 EEKPSRAVAIRIFIGIGSPCYPAQVYGDRRFRWRKYLHLSFHALVGLATEBLQ 1205

```

```

DB 1524 DNTFTGRMIARIFHIGIMSPNYPVAVTGR-CRFRHVVQVDFNRILHLNMAIRK 1576
RESULT 13
AAU34821
ID AAU34821 standard; protein; 610 AA.
XX
AC AAU34821;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #402.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
XX
PR 23-MAY-2000; 2000US-0206848P.
XX
PR 26-MAY-2000; 2000US-0207727P.
XX
PR 23-OCT-2000; 2000US-0242578P.
XX
PR 27-NOV-2000; 2000US-0253625P.
XX
PR 22-DEC-2000; 2000US-0257931P.
XX
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
XX
DR N-PSDB; AAS52680.
XX
PT New polynucleotides for the identification and development of
PT antibiotics; comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 10414; 511bp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 610 AA;
XX
Query Match 9.8%; Score 631.5; DB 4; Length 610;
Best Local Similarity 37.6%; Pred. No. 3,6e-37;
Matches 164; Conservative 63; Mismatches 134; Indels 75; Gaps 15;
QY 474 EQLGHOAFRPGQERAVNRILSGISTLVLPAGAKSLCYQLPALLYSRRSPCLTLVSP 533

```


AC ABU28286;
 XX
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by prokaryotic essential gene #13813.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 OS Enterobacter cloacae.
 XX
 FN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PI 21-MAR-2002; 2002WMO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zykkind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 DR N-PSDB; ACA32156.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 56210; 1766bp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SO Sequence 609 AA;

Query Match 9.6%; Score 619; DB 6; Length 609;
 Best Local Similarity 35.4%; Pred. No. 2.9e-36;
 Matches 169; Conservative 73; Mismatches 152; Indels 84; Gaps 18;

QY 474 EOLGHOAFRPGQERAVNKLISGISTLLVLTPTGAKSKLCYQPLALLYSRRSPCLTLVSP 533
 DB 19 ETPGQGRPGQERTLIETVLEGRDCLVMPFGGKSLCYQVPALVING----LTVVSP 74
 QY 534 LSLMDQVSG-LPPLCLRAACHSGMTRKQRESYVLOKTRAAQVHYMLTPALVAGAGLP 592
 DB 75 ISLMKDQVDQLANGVAAACINSTQTRQOQEVAGCCTGQIRLMYLAPEELMDNFIDH 134
 QY 593 AAQLPVVAFACIDEAHCLSGSHNRPCTYLVCKVLRERMGVHCEGLTATATRTASDV 652
 DB 135 LAHNPVLLA-VDEAHCTSGMHDPREYALGO-LKQRFELPMLTATADDTTRDI 192
 QY 653 AQHLAAVEEPDLHGPAPVPTMHLV-SMDRDTQALLTLQGRFQNLDSI----- 703
 DB 193 VRLDG-----LNDP-----YIQVSSFDPNIRYMLM-----EKFKPLDQLLRVVGQR 235
 QY 704 ----IYCNREDTERIALIRTCLHAAMVPGSGRAKTTAEVYHAGMCSRRRRVQRA 759
 DB 236 GKSGLIYCNRSRAKYEDTAARLQN-----RGFSAAAYHAGLEHHIRADVOEK 281
 QY 760 FMQGLRVVAVTVARGMGLDRPDVRAVLHLGLPSEFESYVQAVGRAGRDGPACHLFL 818
 DB 282 FORDDLQIVAVTVARGMGLDRPDVRAVLHLGLPSEFESYVQAVGRAGRDGPACHLFL 341
 QY 819 -----OPGEDLRELRRH---VHADSTDFLAVRLV-----QVFPACTC 855
 DB 342 PADMAKLRCLEKRGQG-LQDIERHKLMMGAFNAQTCRLVLVNFGEGRQPCQNC 400
 QY 856 -TCTRPSEGRGAVGERPVVK-YPPQAEQLSH-----QAAPEFRVCMGHERALPT 906
 DB 401 DICLPPKQYDGLMDARALSTIYVNRQFGMGVVEVLRGANNRIRIDMGDK-LPV 457

Search completed: December 27, 2005, 21:48:06
 Job time : 224.8 secs

THIS PAGE BLANK (USPTO)